9574–9591 *Nucleic Acids Research, 2011, Vol. 39, No. 22* *Published online 2 September 2011*
doi:10.1093/nar/gkr672

Role of sequence encoded *κB* DNA geometry in gene regulation by Dorsal

Nirotpal Mrinal[1,2,*], Archana Tomar[1] and Javaregowda Nagaraju[1,*]

[1]Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Nampally, Hyderabad 500001 and [2]Molecular Biology & Genetics Laboratory, Faculty of Life Sciences and Biotechnology, South Asian University, JNU Campus, New Delhi 110067, India

Received October 26, 2010; Revised July 28, 2011; Accepted July 29, 2011

ABSTRACT

Many proteins of the Rel family can act as both transcriptional activators and repressors. However, mechanism that discerns the 'activator/repressor' functions of Rel-proteins such as Dorsal (*Drosophila* homologue of mammalian NFκB) is not understood. Using genomic, biophysical and biochemical approaches, we demonstrate that the underlying principle of this functional specificity lies in the 'sequence-encoded structure' of the *κB*-DNA. We show that Dorsal-binding motifs exist in distinct activator and repressor conformations. Molecular dynamics of DNA-Dorsal complexes revealed that repressor *κB*-motifs typically have A-tract and flexible conformation that facilitates interaction with co-repressors. Deformable structure of repressor motifs, is due to changes in the hydrogen bonding in A:T pair in the 'A-tract' core. The sixth nucleotide in the nonameric *κB*-motif, 'A' (A_6) in the repressor motifs and 'T' (T_6) in the activator motifs, is critical to confer this functional specificity as $A_6 \rightarrow T_6$ mutation transformed flexible repressor conformation into a rigid activator conformation. These results highlight that 'sequence encoded *κB* DNA-geometry' regulates gene expression by exerting allosteric effect on binding of Rel proteins which in turn regulates interaction with co-regulators. Further, we identified and characterized putative repressor motifs in Dl-target genes, which can potentially aid in functional annotation of Dorsal gene regulatory network.

INTRODUCTION

Transcription factors (TFs) are DNA-binding proteins that bind to cognate DNA motifs with sequence specificity to regulate gene expression. They are capable of homing on the correct binding site out of a vast number of potential sites scattered in the genome, by virtue of having a surface that is chemically complementary to that of the DNA motif. The current model of DNA motif recognition by a TF is based on the sequence dependent read out of H-bond donors and acceptors in the major groove (1). Although efforts have been made to understand the characteristic chemical signature of nucleotides in the major groove, there is as yet no protein–DNA code for base recognition (2,3).

There are different ways in which a protein surface can take a structure that is 'chemically' complementary to DNA of a particular sequence e.g. through different types of DNA-contacting protein folds, flexibility of protein side chains, TF-induced structural changes in DNA, etc. Despite the lack of a well-defined set of rules governing sequence recognition, some principles and common themes in DNA–protein interactions have emerged (4,5). The current paradigm is that sequence specificity in DNA–protein interaction comes from precise near collinear apposition of donor and acceptor groups leading to formation of hydrogen bonds between the protein and the DNA (3).

The grooves of DNA are rich in hydrogen bond-forming functional groups because of which substantial number of intermolecular hydrogen bonds are observed in protein–DNA complexes (6,7). Traditionally, contacts between DNA and protein have been explained in terms of direct hydrogen bonds, water-mediated hydrogen bonds, van der Waal interactions, electrostatic and, hydrophobic contacts (8–10). The role of water-mediated hydrogen bond was revealed in the crystal structure of the *trp* repressor–operator-specific complex in which several water-mediated contacts between protein and DNA were observed but no hydrogen bond to the bases (11,12). Furthermore, structure of phage 434 repressor–operator complexes revealed role of non-contacted bases in binding, implying a role for base sequence-induced DNA structure (13).

*To whom correspondence should be addressed. Tel: +91 11 2674 1471; Fax: +91 11 2674 1741; Email: nmrinal@sau.ac.in
Correspondence may also be addressed to Javaregowda Nagaraju. Tel: +91 40 2474 9342; Fax: + 91 40 24749448; Email: jnagaraju@cdfd.org.in

Later, crystal structures of DNA with catabolite gene activator protein, IHF and TATA-binding protein (14–19) illustrated the role of sequence-dependent structural adaptation in DNA. These studies laid the foundations of a structural code for DNA recognition (20).

The role of DNA structure and its sequence-dependent structural adaptation to facilitate protein binding has only recently begun to be understood (21,22). Apart from sequence specificity, other factors that determine the interaction between a TF and its cognate binding motif are poorly understood, e.g. interactions with phosphate backbones in DNA–protein complexes are also common but they are relatively less studied (5). It is now accepted that TFs recognize a sequence-based DNA shape which is appropriately called as shape readout. Shape readout is an important concept as it can explain specificity in DNA–protein interactions at higher resolution (20–22). It includes both global (e.g. DNA bending) as well as local shape (e.g. major and minor grooves) recognitions. These local and global structural readouts are sequence dependent, e.g. A-T rich sequence takes typical B-DNA form while GC-rich motif assumes A-form (23–26). The structure assumed by A-tract DNA (stretch of >4 A residues) is bent with a narrow minor groove, and is sometimes called B′-DNA (27–29).

The critical differences in structural changes in different DNA forms is best revealed in the electrostatic potential differences as it has been shown that minor grooves locally enhance negative electrostatic potential (30,31). Hence any change in minor groove shape is likely to change the electrostatic potential which in turn will affect DNA–protein interactions. In such a scenario, can the logic of transcriptional regulation be reliably understood merely by the matching of consensus TF sites from the genomic sequences? For example, a study of SRY and SOX family proteins in mouse revealed that each of these TFs has the ability to identify two distinct binding sites: one primary and the other secondary (32). Thus, a major limitation in understanding transcriptional regulation is the rarity of determined specificities for a majority of the encoded TFs. Recent studies highlighting the role of minor groove shapes in imparting specificity of TFs binding to their target genes indicate the existence of more intricate programming of gene regulation than previously thought (31–33).

Binding of TFs to *cis*-elements leads to activation/repression of gene transcription, and this has led to classification of TFs as repressors or activators e.g. KRAB family of Zinc-finger TFs are known transcriptional repressors (34). An example of transcriptional activator is CF2 (Chorion Factor 2), which possesses a transcriptional activation domain, but, its role as transcriptional repressor has also been documented (35). These studies suggest that gene activation/repression by a TF is context dependent and thus they highlight the complexity in assigning a TF as an activator or as a repressor. There are certain TFs, called bifunctional transcriptional factors, which can activate as well as repress gene transcription. Many developmentally important TFs fall under this category, e.g. Dorsal (Dl), a morphogen, activates as well as represses target genes during embryonic dorso-ventral patterning in *Drosophila* (36,37). It is not known that how Dl recognizes 'the addresses' of the target genes as activators or repressors. In other words, how does Dl decide which genes to activate and which to repress?

Dl, like other proteins of the Rel family, binds to a 'loosely' conserved DNA sequence (GGGRRYYCCC) called as *κB*-motif (37). Predicting a functional *κB*-motif is particularly difficult because of the sequence heterogeneity. Furthermore, it has not been ascertained whether the *κB*-motif sequence plays any role in activation or repression of target gene transcription by Dl. It has been shown that a single base change in the *κB*-motif can determine the co-factor specificity of NFκB dimers thus highlighting the role of *κB* sequence in Rel-mediated gene regulation (38). Recently, we demonstrated 'sequence specificity' in Dl interaction with its co-regulator AP1, as we found that the Dl–AP1 complex was recruited on an atypical *AGAAA AACA* motif but not on canonical *GGGAATTCC* motif in the same promoter (39). This raised the question as to how Dl-binding motif sequence decides co-regulator specificity.

In the study reported here, we investigated 'nucleotide signatures' in Dl-binding motifs with respect to their activator/repressor functions. We show that the ability of Dl to activate or repress target genes depends on the geometry of *κB*–DNA which in turn depends on its sequence. The genome-wide analysis of different *κB*-motifs in *Drosophila* revealed distinct bias in the sequence of *κB*-motifs present in genes repressed by Dl as compared to those activated by it. We have also investigated the effect of sequence change on DNA structure by molecular dynamics studies, and show that different activator motifs, in spite of sequence differences, have comparable major groove geometries. Our findings indicate that the structure of the *κB*-DNA backbone is an important factor that determines not only the ability of Dl to bind the *κB*-motif but also its ability to interact with cofactors.

MATERIALS AND METHODS

Homology modelling

The protein–DNA system was modelled using default parameters of Modeller 9v2 (40). Since the crystal structure of Rel proteins of *Drosophila* or other insects is not known, we used chicken c-Rel bound to *κB*-motif of CD28 promoter (PDB code-1GJI) as template for modelling RHD of Dl bound to DNA (41). The RHD of chicken c-Rel was closest to the RHD of the Dl as they share 48% homology and their DNA-binding residues are completely conserved (Supplementary Figures S1 and S2). Furthermore, the sequence of CD28 responsive element *AGAAATTCC*, is close to that of the $dl\kappa B^P$ motif (*AGA AAAACA*) and hence it was chosen as the appropriate DNA template for homology modelling of the $dl\kappa B^P$ motif.

To generate the structure of the $dl\kappa B^P$ DNA, the sixth and seventh 'T'and the nineth 'C' in the crystal structure of *AGAAATTCC* was replaced with 'A'. While replacing the original base, nitrogen atom bonded to the sugar and the two base carbon atoms bonded to this nitrogen were aligned with the corresponding atoms of the new base so

that the sugar base bond and phosphate backbone remain unchanged. In other words, the base substitution was done without affecting the backbone of DNA.

To model the DNA in the protein–DNA complex, the DNA atoms were defined as 'HETATM' and the residues as 'BLK' (Modeller 9v2) in the template. These residues are restrained more or less as rigid bodies to retain the conformation of the equivalent residues in the template.

The mutations were introduced by keeping the orientation of the structure with reference to the first base pair by using the frame_mol utility of X3DNA package (42). Furthermore, using the rebuild facility the co-ordinates of the atoms were regenerated and the nitrogenous base co-ordinates were manually replaced in the structure, keeping the phosphate backbone constant. DNA structure generated after mutation was aligned with template DNA and no significant distortion in the backbone was found (RMSD, 0.0008).

Molecular dynamics

All simulations were performed using the molecular dynamics program NAMD2 (43) and CHARMM22 force field. The resulting structure was then subjected to 2000 steps of initial minimization to remove bad contacts and reduce the strain in the system. The complex was then immersed in the centre of a box of radius 10 Å filled with TIP3P water. In order to obtain the proper geometry for each water molecule, all oxygens were held fixed and 1000 steps of energy minimization of the bond and angle energies was performed. All atoms were then relaxed, and the entire system was equilibrated at 300 K for 30 ps. Water molecules closer than 1.8 Å to the protein–DNA system were removed. The resulting system was further subjected to 5 ps equilibration during which the protein–DNA backbone of the system was put into constraints.

To achieve electro-neutrality for the system, 30 Na^+ and 15 Cl^- were added by removing 45 water molecules, located >9 Å apart from each other and 5 Å apart from the protein or DNA atoms. The system was further equilibrated for 15 ps without any constraints. This system contained 1,180,85 atoms, and was finally subjected to molecular dynamics simulation for 10 ns. Constant temperature was maintained at 300 K using Langevin dynamics with a damping coefficient of 1 ps^{-1}. Short-range non-bonded terms were evaluated every step using a 10 Å cut-off for van der Waals interactions. An integrated time-step of 1fs (DCD freq 500) was used. The system was simulated with periodic boundary conditions and full electrostatics computed using the Particle Mesh Ewald (PME) method with a $120 \times 125 \times 108$ point grid. The resulting trajectories were analysed by VMD (44). The conformational changes of the DNA during dynamics were evaluated using the X3DNA program (42). Major groove and other DNA local structural parameters were also analysed using the same X3DNA package. All simulations were run on SunGrid Engine running the Red Hat Enterprise operating system for AMD architecture.

Transfection, luciferase reporter assay

Protocols for cell transfection and luciferase assay were as described by Mrinal and Nagaraju (39).

Electrophoretic Mobility Shift Assay (EMSA)

The nuclear extracts were prepared as mentioned previously (39). A total of 100 ng of double-stranded oligo was labelled with 3 µl of [γ-^{32}P] ATP and 1 µl of polynucleotide kinase in 1 µl PNK buffer (New England BioLabs) for 1 h at 37°C. The labelled DNA was purified on a G50 column. DNA-binding reactions were done in 5 mM Tris–HCl, pH 7.9; 12 mM HEPES; 50 mM KCl; 3 mM EDTA; 1 mM DTT; 5 mg/ml BSA; 10% glycerol; 0.1 mg/ml poly(dI-dC). Nuclear extracts prepared from LPS+PGN-treated cells were pre-incubated in this buffer in the presence/absence of 40-fold excess of unlabelled oligonucleotide at room temperature for 15 min. Afterwards, 50–100 pg of the labelled oligonucleotide was added to the reaction mix, which was then incubated for another 15 min at 25°C. The binding reactions were analysed by electrophoresis on 6% native polyacrylamide gels.

RESULTS

κB-motif has variable sequence but distinct organization

Rel proteins are multidomain proteins. However, their characteristic feature is the presence of β-sheet sandwich immunoglobulin fold. Such DNA-binding proteins typically use loops for interaction with DNA (21,41). Interaction of different Rel proteins with cognate *κB*-motifs is also mediated through a series of loops. While, RHD, the DNA-binding domain of Rel proteins is highly conserved, *κB*-motifs, which interact with RHD of the Rel proteins, probably represent the most heterogeneous DNA-binding motifs. It is currently not known why such heterogeneity may have evolved (Figure 1A). In spite of sequence variations, the nucleotides in *κB*-motifs are arranged in a dyad fashion with two half-sites joined by a central hinge (41) (Figure 1A). Here, we explored the functional significance of sequence variations in *κB*-motifs.

Terminal nucleotides G in the first half-site and C in the second half-site are critical for DNA–proteins interactions and hence are more conserved (Figure 1A and B). While nucleotides in the first and second half-sites are always protein contacted the nucleotides 4–6, which form the core of the *κB*-motifs, usually, do not interact with the Rel proteins directly (Figure 1B) and hence they would probably have experienced less selection pressure compared to the terminal nucleotides which form direct interactions with the Rel proteins. In the absence of selection pressure, the probability of occurrence of any of the 4 nucleotides would have been expected to be equal in the core region. However, A/T rich core region shows position-specific preference for T at the sixth and seventh positions and A at the fourth and fifth positions (Figure 1B–D). An earlier SELEX study performed with the three *Drosophila* Rel proteins had also suggested

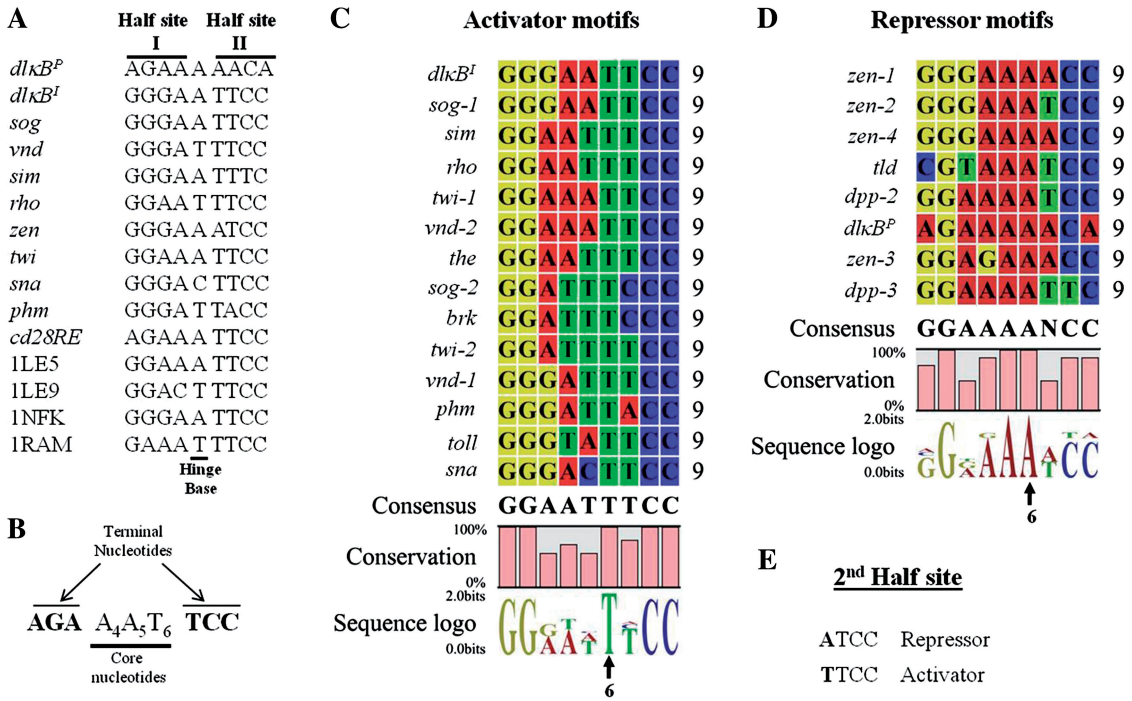

Figure 1. Sequence bias in the composition of the activator and repressor *κB*-motifs. (**A**) The second half-site of the κB-sequence dyad is more conserved than the first half-site. Hinge nucleotide (underlined) of the dyad is usually A or T. (**B**) Only the bases at the termini (shown in bold) form protein contacts while the core bases (underlined) usually do not form hydrogen bonds with proteins. (**C** and **D**) Repressor motifs have 'A-tract' which is lacking in the activator motifs as revealed by Weblogo consensus. The sixth base in the repressor *κB*-motifs is always 'A'. For the repressor motif consensus prediction, the *κB*-motifs present in Dl repressor target genes *dpp* and *zen* were taken into consideration, however only one motif in each gene is functional. Thus all functional repressor motifs have A-tract and the sixth base is always 'A'. (**E**) First half-site consensus is same for both activator and repressor motifs but not the second half site. First base of the second half-site (bold) is A in the repressor motifs and T in the activator motifs.

position-specific preference for A/T in the core region (45). Here, we studied the role and significance of position specific preference for nucleotide, in a *κB*-motif sequence on DNA–Rel interactions in general and DNA-Dl interaction in particular.

Dl can discern activator and repressor *κB*-motifs

Dl, a transcriptional regulator and morphogen, binds GGGYYYYCCC consensus motif (46–48). We aligned *κB*-motifs of Dl-activated and repressed genes separately. Only, the established Dl-target genes were chosen for the analysis. *dlκB^I* and *dlκB^P* motifs are newly identified Dl-binding motifs present in *dl* gene (39). *dlκB^P* is an atypical *κB*-motif (*AGAAAAACA*) and lacks 'T' (Figure 1D). From this comparison, it is evident that the repressor motifs have preference for A to T as a result of which repressor motifs, but not the activator motifs, appear to have A-tract (Figure 1C and D). The sequence differences between activator and repressor motifs are more obvious in their second half-site consensus, which is 'ATCC' for repressor and 'TTCC' for the activator motifs (Figure 1C–E and Supplementary Figure S3). Interestingly, the hinge base (fifth base) in the repressor motifs is always 'A' whereas no such sequence preference is observed for the activator motifs (Figure 1C and D). However, the most noticeable distinction was observed at the sixth base position. It is evident that 'T' at sixth position is present in all activator motifs whereas all repressor motifs have 'A' at the corresponding position (Figure 1C–E).

A single base change can transform an activator *κB*-motif into a repressor motif

Given the distinction of A or T at the sixth position in the activator or the repressor motif respectively, we investigated whether this sequence difference could serve to distinguish between them. We selected two enhancer (*dlκB^I* and *phm*) and two repressor *κB*-motifs (*dlκB^P* and *zen*) for functional comparison by luciferase assay (Figure 2A). *phm* (*GGGATTACC*) and *zen* (*GGGAAATCC*) regulate embryonic development and are known Dl-target genes. While *dlκB^I* is a typical *κB*-motif (*GGGAATTCC*) and acts as an enhancer, atypical *dlκB^P* (*AGAAAAACA*) acts as a repressor (39). For the functional assay, these motifs were placed at the *dlκB^P* position in the *dorsal* promoter construct with a luciferase reporter (39).

For the functional analysis, we introduced an $A_6 \rightarrow T_6$ mutation in the respective repressor motifs, and $T_6 \rightarrow A_6$ mutation in the two activator motifs. We found that activator function of *phm* and *dlκB^I* motifs was switched to that of repressor by $T_6 \rightarrow A_6$ mutation in the respective Dl-binding motifs of *phm* and *dlκB^I* (Figure 2B). Reciprocally, the repressor *zen* and *dlκB^P* motifs became inducible upon $A_6 \rightarrow T_6$ mutation (Figure 2C).

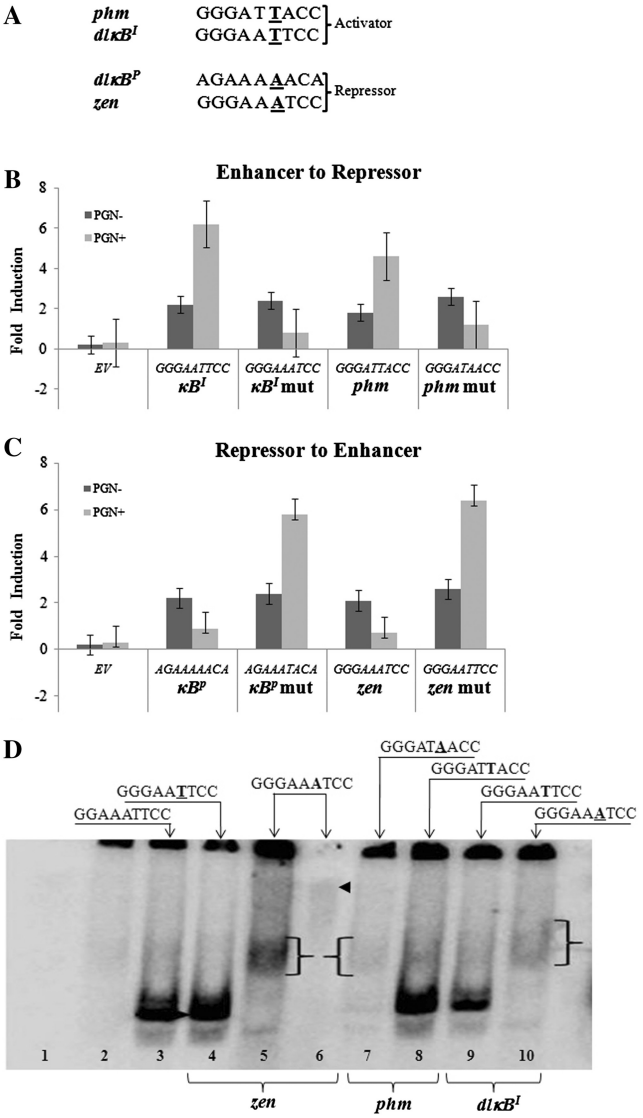

Figure 2. Enhancer or repressor activity of *κB*-motifs is encoded in its sequence. (**A**) Two activator motifs $dl\kappa B^{I}$ and *phm* and two repressor motifs $dl\kappa B^{P}$ and *zen* were used to decipher the sequence code of gene activation or repression. The sixth nucleotide is bold and underlined. (**B**) $T_6 \rightarrow A_6$ mutation transforms enhancer motif into repressor. (**C**) Reciprocal mutation, $A_6 \rightarrow T_6$ in the repressor motif confers it enhancer activity. (**D**) Dorsal interaction with different *κB*-motifs is sequence-dependent as seen in EMSA. Activator motifs (lanes 3, 8 and 9) form smaller DNA–protein complex while repressor motifs (lanes 5, 7 and 10) form larger complexes indicating the presence of additional proteins. The *zen*–repressor complex (as shown by brackets) (lane 5) was supershifted with anti-Dl antibody (lane 6) confirming the presence of Dl in the complex (indicated by arrow head). $A_6 \rightarrow T_6$ mutant of the *zen* motif retards smaller complex (lane 4) similar to control enhancer motif *twi* (lane 3) suggesting that Dl interaction with the co-repressor is lost due to $A_6 \rightarrow T_6$ mutation resulting in a smaller complex. Activator motifs phm (lane 8) and $dl\kappa B^{I}$ (lane 9) retard small complexes. However, their $T_6 \rightarrow A_6$ mutant probes retard larger complexes (as shown in brackets) (lanes 7 and 10). The probe sequences are given at the top of the lane. The sixth nucleotide is shown in bold while corresponding mutant nucleotides are underlined. Lanes: 1—cold competition with the specific oligo probe, 2—mutant oligo, 3—*twi* oligo, 4—T_6-mutant of *zen*, 5—*zen*, 6—supershift with Dl antibody, 7—A_6-mutant of *phm*, 8—*phm*, 9—$dl\kappa B^{I}$ probe, 10—A_6-mutant of $dl\kappa B^{I}$.

We obtained complete transformation of enhancer motifs into repressors by mutating sixth 'T' to 'A' and *vice versa* by mutating sixth 'A' to 'T'. However, no change in gene expression was observed for any of the four *κB*-motifs upon $A_7 \rightarrow T_7$ or $T_7 \rightarrow A_7$ mutations. In fact, $A_7 \rightarrow T_7$ mutation in the *phm κB*-motif led to further enhancement of gene expression suggesting that the *GGG ATTTCC* motif is a strong activator compared to *GGGATTACC* motif (Supplementary Figure S4). These observations are consistent with those of Muroi *et al.* (49) which showed that nucleotide substitutions at the seventh position in a *κB*-motif are tolerated but not those at the sixth position.

These results also suggest that position-specific A or T bias at the sixth position in the enhancer or repressor motifs does carry functional implications. In other words distribution of A or T at in the *κB*-motif is not a random feature.

A_6 is critical for Dl interaction with co-repressor

Reversal of Dl-mediated gene expression (repressor to enhancer and vice versa) by single nucleotide change in the *κB*-motifs was intriguing, and we set out to study how nucleotide A or T at the sixth position in the *κB*-motif can specify activator or repressor function of Dl. Repression of gene expression by Dl requires interaction with co-regulator proteins like Groucho and dCTBP (50). Ventral repression of *zen* is dependent on interaction of Dl with an uncharacterized protein that binds to the neighbouring AT-rich motifs on the *zen* repressor element (51). These studies have led to the hypothesis that transcriptional repression by Dl is co-regulator dependent (51).

Recently, we identified AP1 as co-repressor of Dl and showed that Dl-AP1 complex binds to $dl\kappa B^{P}$ motif (which has A as the sixth nucleotide) but not to $dl\kappa B^{I}$ motif (which has T as the sixth nucleotide), although both motifs are functional and present in the same *dl* gene. $dl\kappa B^{P}$ motif and the AP1-binding cluster in the *dl* promoter are in close proximity. However, when (i) the $dl\kappa B^{P}$ motif was replaced with $dl\kappa B^{I}$ motif or, (ii) the two motifs were swapped, there was no interaction between AP1 and Dl bound to $dl\kappa B^{I}$ motif (39). These findings highlighted the role of DNA sequence not only in Dl binding but also in the interaction of Dl with its co-regulator.

Thus, our data suggested that transcriptional repression by Dl requires (i) A_6-*κB*-motif, and (ii) a gene specific co-repressor. We hypothesized that possibly A_6 in the repressor *κB*-motif is essential for the assembly of Dl co-regulator complex, which in turn represses the target gene (Figure 2). To decipher putative modulation of Dl–co-regulator complex binding to a *κB*-motif by its nucleotide sequence, we performed EMSA. Our assessment was that if sequence of the *κB*-motif-regulated Dl interaction with its co-regulator, then biochemically distinct Dl–DNA complexes should be seen in EMSA. Role of A_6-*κB* in Dl-mediated repression was tested for Dl-repressed gene *zen*. With *zen* motif as probe, a larger Dl–DNA complex was indeed observed compared to that with a T_6-mutant of the same probe (Figure 2D, lanes 4–6), suggesting that the single nucleotide change ($A_6 \rightarrow T_6$) in the *zen* motif

results in loss of interaction of Dl with its co-regulator although it does not affect binding of Dl per se. On the other hand, *phm* and *dlκB*^{I}, both activator motifs, yielded smaller complexes of similar size with Dl (Figure 2D, lanes 7–10). Mutation of $T_6 \rightarrow A_6$ in these motifs resulted in larger DNA–protein complexes, indicating the presence of additional proteins in them (Figure 2D, lanes 7 and 10). Thus the EMSA results suggest that Dl interaction with co-regulator is possibly modulated by the sequence of the *κB*-motif. These findings led us to postulate that the function of Dl as an activator or repressor is probably encoded in the sequence of its binding motif.

Enhancer *dlκB*^{I} and repressor *dlκB*^{P} motifs have different structures

DNA–protein interactions always occur through the major and/or minor groove of the cognate DNA-binding motifs. To elucidate how single base change in a *κB*-motif completely reverses its transcriptional behaviour and its interaction with co-regulatory proteins, we undertook a biophysical approach. We generated structures of different *κB*-motifs, taking CD28 Rel-binding motif (*AGAAATTCC*) as a template, by homology modelling followed by molecular simulation (Figure 3A and Supplementary Figure S5 and details therein) (41).

The model of Dl–DNA complex showed binding of the Dl monomers through the major groove and hence, we compared geometry of major grooves of the enhancer motif *dlκB*^{I} and the repressor motif *dlκB*^{P} (Figure 3B–E). It is evident that the repressor *dlκB*^{P} motif has a significantly reduced major groove at the fourth and the fifth base positions with the maximum reduction for the latter, the hinge position (Figure 3B and D). On the other hand there were no significant structural differences between the major grooves of the two enhancer motifs *dlκB*^{I} and CD28-*κB* (Supplementary Figure S5).

Major groove geometry of the CD28-*κB*-motif, as inferred from the molecular dynamics studies, was not

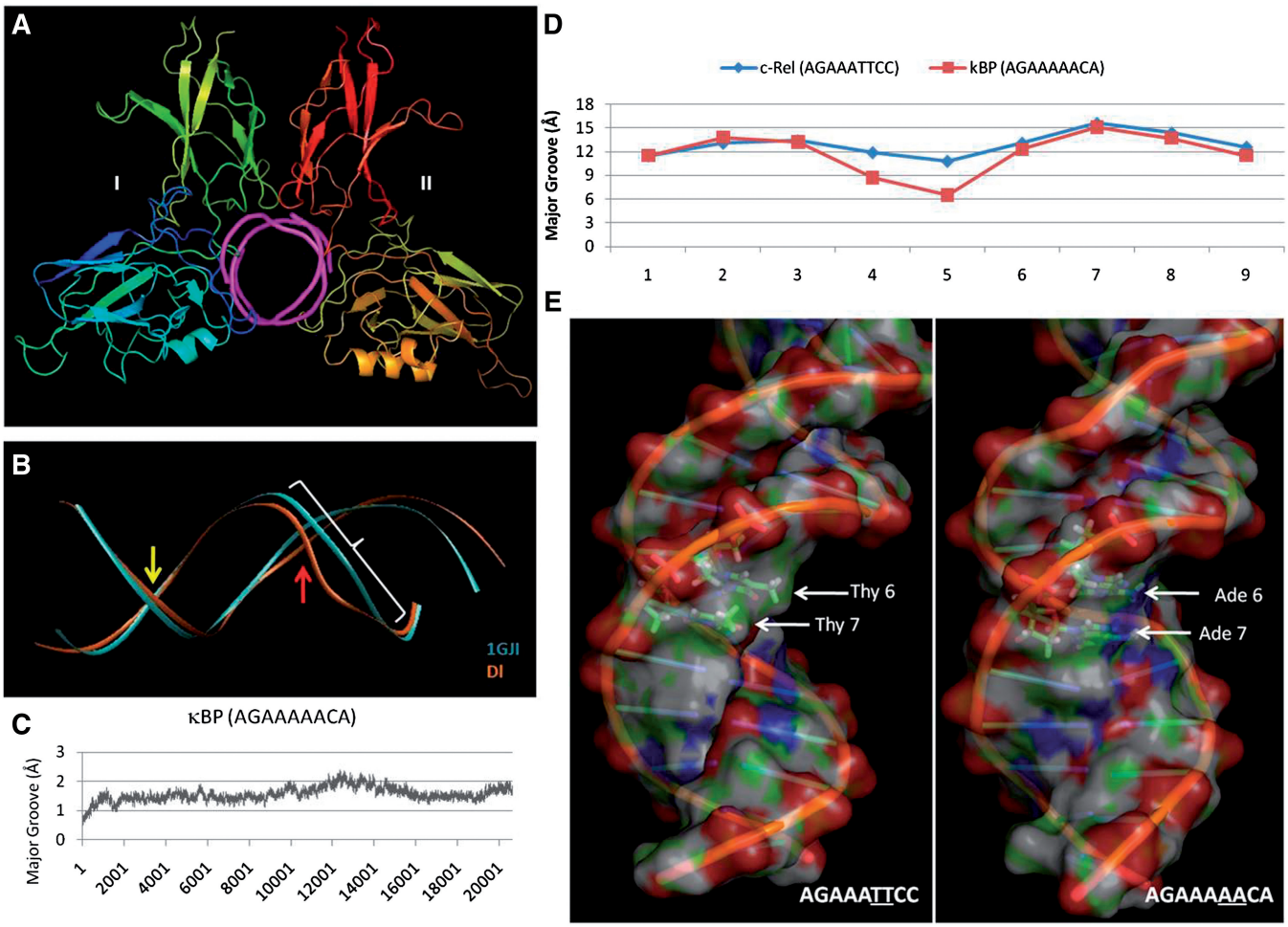

Figure 3. *AGAAATTCC* and *AGAAAAACA* motifs have different structures. (**A**) Structure of **d**orsal homodimer bound to enhancer motif (*AGAAATTCC*) generated after homology model**l**ing is shown. (**B**) Repressor *dlκB*^{P} motif, *AGAAAAACA* shows kink conformation compared to enhancer *AGAAATTCC* motif of CD28RE (1GJI). (**C**) RMSD graph shows stable simulation of simulation *dlκB*^{P} DNA backbone (*AGAAAAACA*) during 10 ns (20 000 frames) molecular dynamics. (**D**) Major groove of *dlκB*^{P} motif shows reduced geometry at the fourth and the fifth base. Sharp dip at the fifth nucleotide position is striking. (**E**) The space filling model of the enhancer *AGAAATTCC* and repressor *AGAAAAACA* motifs shows change in the major groove conformation due to mutation of T_6T_7 to A_6A_7 as shown here. Sharp turn at the fifth nucleotide position and extremely reduced major groove can be seen. Presence of 'T' or 'A' at the sixth and seventh positions are indicated.

different from that seen in its crystal structure, indicating that no structural changes were introduced during the simulation process (Supplementary Figure S5). Furthermore, while incorporating the mutations in the structure of the template DNA, the bases alone were replaced without altering the phosphate backbone. Taking these points into consideration, therefore, we suggest that the unique geometry of the $dl\kappa B^P$ DNA backbone generated following the same simulation protocol reflects its structural property and that it is not a simulation-generated structural artefact (Figure 3).

Recent studies have suggested that many major groove-binding TFs also interact with the minor groove, e.g. homeodomain proteins are major groove binders but they also extend their basic amino acids into the minor groove to impart specificity (31). Arginine is the most abundant residue that inserts into minor grooves, and lysines are also observed in such regions although less favoured (33). Minor groove width is dependent on sequence composition (27–29). Hence, we compared the minor groove of the $dl\kappa B^P$ and $dl\kappa B^I$ motifs. The minor groove of the repressor $dl\kappa B^P$ motif is straight, which is typical of A-tract motifs (29). We did not find any significant decrease in the minor groove width of the $dl\kappa B^P$ motif compared to that of the activator CD28-κB-motif (Figure 4A). Next, we checked whether narrowing of the major groove led to interaction of basic amino acids such as arginine and lysine of Dl in the minor groove of the $dl\kappa B^P$ motif. Rel proteins bind the DNA through the major groove. Whether Dl, which is a Rel homologue, binds only through the major groove or also involves minor groove was checked for the activator motif $dl\kappa B^I$ as well as for the repressor motif $dl\kappa B^P$. We performed competition experiments with the major and minor groove binding drugs and resolved the complexes by EMSA. Dl binding was affected only when competition was performed with the major groove binding molecule but not with the minor groove-binding drug (Figure 4B and C). These results provide biochemical evidence that minor groove is not involved in the binding of Dl–AP1 complex to the $dl\kappa B^P$ motif (Figure 4C). These results also corroborate our structural modelling data (Figure 3A–E). Usually reduction in major groove width is associated with compensatory increase in the minor groove. However, in case of $dl\kappa B^P$ motif, narrowing of the major groove at the fifth position was not associated with increase in the minor groove width (Figures 3D and 4A). Since, geometry of the minor groove remained unchanged at the fifth base

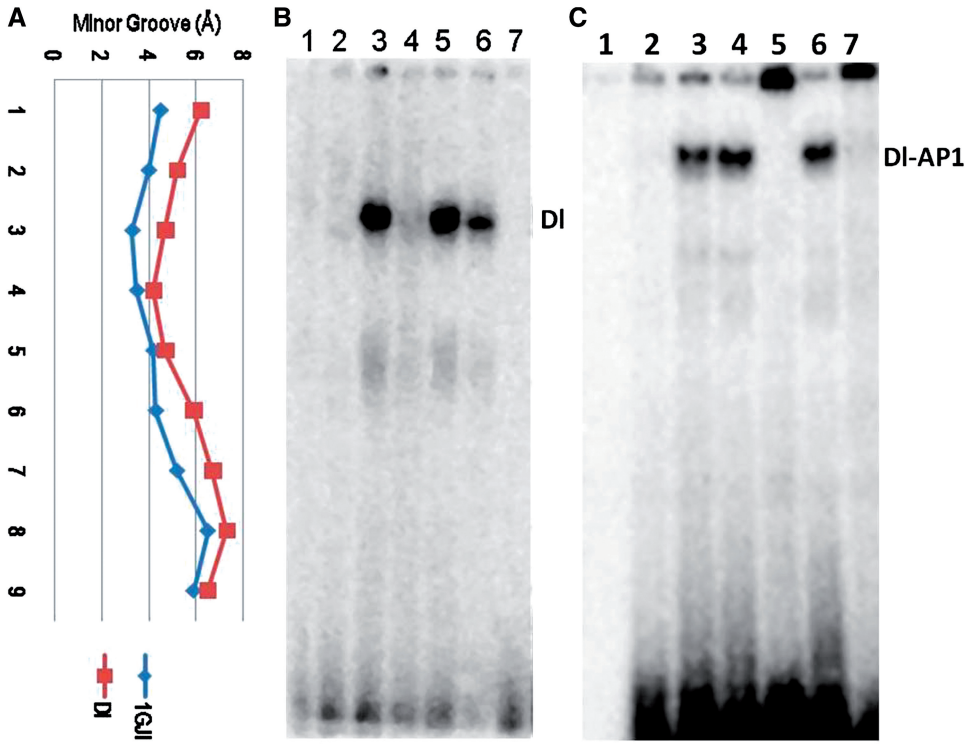

Figure 4. Dl binding to the $dl\kappa B^P$ motif does not involve minor groove interactions. (**A**) Minor groove of $dl\kappa B^P$ motif (Dl) is uniform and straight. Minor groove width of $dl\kappa B^P$ motif is bigger compared to that of cd28-κB-motif (1GJI) at most of the nucleotide positions except the fifth nucleotide. (**B**) The Dl–DNA model showed the two Dl monomers bound to major groove of κB-DNA (Figure 3A and Supplementary Figure S1). Whether or not Dl interacts with the major groove of κB^I-motif was examined by performing competition experiments with major and minor grooves binding drugs and resolved by EMSA. Dorsal binding was lost upon competition with major groove binding drug, methyl green, but not upon competition with minor groove binding drug Hoechest. This suggests that there are no Dl interactions in the minor groove of the κB-motif. (Lanes, 1—mutant probe, 2—cold competition, 3—κB^I motif, 4—Methyl Green competition, —Hoechest competition, 6—non-specific competition, 7—free probe). (**C**) Binding of Dl–AP1 complex on the repressor $dl\kappa B^P$ motif also does not involve interactions with the minor groove. (Lanes, 1—homologous cold competition, 2—mutant probe, 3—κB^P motif probe, 4—Hoechest competition, 5—Methyl Green competition, 6—non-specific competition, 7—free probe).

position which might have constrained the major groove at this position, ultimately resulting in sharp narrowing of the major groove or the formation of the kink in the $dl\kappa B^P$ motif (Figure 3D and E).

Role of κB major groove geometry in Dl–DNA interaction

Though the DNA-binding RHD of Rel proteins is evolutionarily highly conserved, their interaction with their cognate κB-motifs is very specific, partly because orientation of the immunoglobulin folds along the DNA and base contacts are different for different Rel homo/hetero-dimers (21, 41). Specificity in Rel–κB interaction can be accounted for by the DNA contacting loops which in contrast to helices and sheets are not rigid structures. The DNA contacting loops of Rel proteins are not constrained because the side chains of the C-terminal domain of the Rel monomers, which regulate dimer formation, lie along one face of the immunoglobulin sandwich (21). As a result, these loops are free to take different conformations. Furthermore, the role of κB-DNA geometry cannot be overlooked as Rel proteins also interact with the DNA backbone.

Dl, like other Rel proteins, exhibits extensive electrostatic interactions with the phosphate backbone of the core nucleotides apart from forming direct hydrogen bonds with the nitrogenous bases (Supplementary Figure S1). Furthermore, DNA backbone conformation is sequence dependent (21,23–29,31,33). In order to understand the role of 'sequence-encoded backbone geometry' in DNA-Dl interaction, we asked the following questions: (i) How do nucleotide changes affect the DNA backbone geometry? (ii) Do activator and repressor motifs have distinct structures?

Position specific requirement for T_6 in the activator and A_6 in the repressor motifs was intriguing considering that this base is not directly contacted by Dl or chicken c-Rel proteins (Figures 1 and 3) (41). This suggests that Dl is able to discriminate the geometry of an A–T base pair from that of a T–A base pair, which is possible since the carbonyl group of '*T*' and amino group of '*A*', if superimposed in the same plane, are separated by 1.1 Å; hence, A–T and T–A are structurally not similar (52). Consequently, A–T and T–A phosphate backbones would also not be similar, as evident from the backbone geometry of these two base pairs (Figure 3). This may explain the structural differences in the major groove of the $dl\kappa B^I$ and $dl\kappa B^P$ motifs, although both have equal number of A–T pairs (Figures 3D, E and 4A).

To address the question how A–T or T–A distinctly affect the κB-DNA geometry, we sequentially mutated the nucleotides at the fifth, sixth, and seventh positions of the CD28-κB activator motif $AGAA\underline{\mathbf{A}_5}\underline{\mathbf{T}_6}\underline{\mathbf{T}_7}CC$. The effects of these mutations on the DNA geometry ranged from mild to severe (Figure 5 and Supplementary Figure S6). It is also evident that all the motifs that have *T* as the sixth nucleotide exhibit similar major groove geometry and comparable DNA-Dl interactions (Figure 5D–F) when compared to the crystal structure of the c-Rel-cd28κB, DNA–protein complex (Figure 5A). Interestingly, all these motifs also act as activator motifs (Figure 2 and Supplementary Figure S7). However, the major groove geometry of A_{6-} containing κB motifs (Figure 5B and C) is distinct from that of the T_6 motifs (Figure 5D–F). We found that these two motifs have constrained major grooves and form repressor type DNA-Dl complexes (Figure 5B, C and G, lanes 6 and 7) (39).

The most significant change in the major groove geometry was observed for $T_6 \rightarrow A_6$ mutation (Figure 5C; Supplementary Figures S6 and S7). Interestingly this is the same mutation that leads to reversal of activator function to that of repressor, thereby suggesting that A or T at the sixth position is the most important nucleotide in imparting shape to the κB–DNA (Figure 2 and Supplementary Figure S6). Residues R15, R17, E21 and K174 of monomer I of Dl interact with G_2, A_3 of the first half-site whereas R15, R17, E21 and K174 of monomer II form H-bonds with A_7, C_8 and A_9 of the second half-site of the κB-DNA (Supplementary Figure S8). The DNA-binding residues of monomer I are in similar orientation resulting in their normal interaction with G_2, A_3 in both the activator and repressor motifs (Supplementary Figure S8). On the other hand, monomer II interaction with the second half-site nucleotides is different between the activator and repressor motifs, especially interaction of K174 (Figure 5B and C; Supplementary Figure S8). Interestingly, K174 exhibits electrostatic interaction with the phosphate backbone of upper strand of *AGAAAAAC A* and lower strand of *AGAAAATCC* indicating that Dl interaction with different repressor motifs is probably unique (Figure 5B and C).

We have shown above that Dl-co-regulator interaction is sequence specific and co-regulator dependent (Figure 2D). $T_6 \rightarrow A_6$ mutation in the CD28-binding element, which is an activator κB-motif, reduces the major groove at the fifth position which probably facilitates Dl interaction with the co-repressors as seen in EMSA (Figures 2, 3, 5B, C and G, lanes 6 and 7, respectively). The T_6-activator motifs have similar DNA geometry and also form Dl–DNA complexes of similar sizes (Figure 5G, lanes 1–5 and Supplementary Figure S6). Taken together, these data indicate that Dl interaction with the repressor motifs is co-regulator dependent. Since different co-regulators would interact differently with Dl, binding of Dl-repressor complexes on DNA would also be different (Figures 2D and 5G).

O_4–N_6 hydrogen bond of sixth A in repressor κB-motifs imparts flexible major groove geometry

As described above, molecular simulation studies demonstrated that repressor κB-motifs have deformable major grooves. Next, we examined how the A-tract core imparts flexibility to the repressor motifs. We compared DNA local parameters, viz. roll, buckle etc of activator and repressor κB-motifs. Buckle and opening showed much larger differences compared to other parameters (Supplementary Figure S9). Interestingly, buckling increased with increase in number of *A* residues towards the 3′-end (least for 3A motif and maximum for 5A

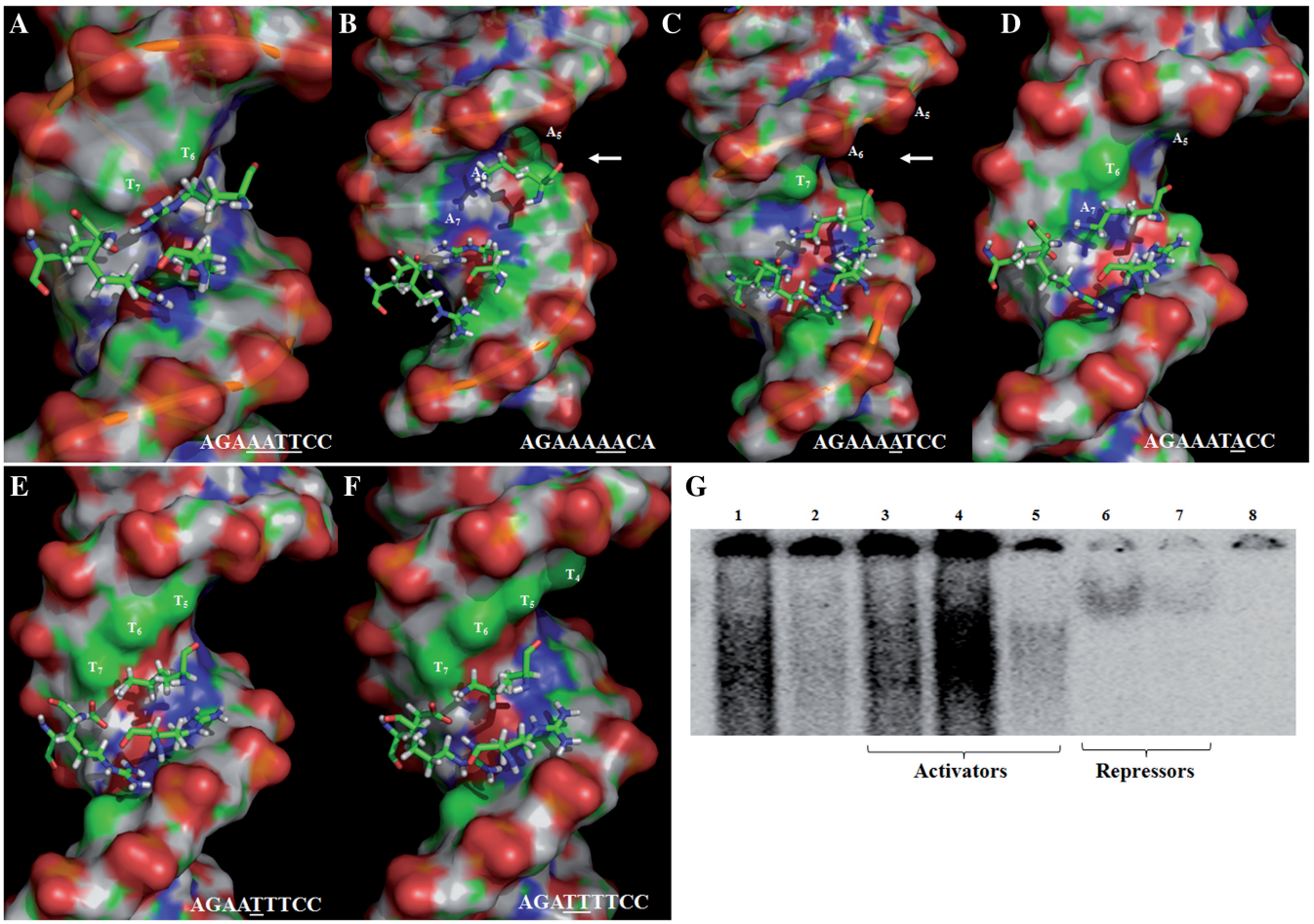

Figure 5. Effect of mutations in the core region of the *κB*-motif, on its geometry. (**A–F**) Effect of the base substitution on Dl interaction with the *κB*-motif was analysed following 10 ns simulation. (A) Interaction of the chicken c-Rel with the cognate DNA is shown for comparison with other mutants. (B–F) The two repressor motifs (B and C) have reduced major groove (indicated by arrow) in comparison to the activator motifs (D–G). DNA-binding residues of Dl (R15, R17, E21 and K174) show similar interactions in all the activator motifs. However, interaction of the same four amino acid residues of Dl with the two repressor motifs (B and C) is different probably because their major grooves are more constrained. The *AGAAAAACA* repressor motif shows sharp bend at the fifth position whereas another repressor motif *AGAAAATCC* shows bending at the seventh position in the second half-site. (D–F) DNA–protein interactions are almost similar in these activator motifs implying that sequence change does not affect overall DNA–protein interactions with activator *κB*-motifs. All structures were generated and visualized using PyMol. (**G**) EMSA was performed to validate if structural changes seen in DNA affect Dl binding or its interaction with co-regulators. The two repressor motifs (lanes 6 and 7) form bigger size complexes while the activator motifs form complex of almost same size though the affinity of Dl binding varies (lanes 1–5). Lanes 1—*AGAAATTCC*, 2—*AGAAATACC* ($T_7 \rightarrow A_7$), 3—*AGAATTTCC* ($A_5 \rightarrow T_5$), 4—*AGATATTCC* ($A_4 \rightarrow T_4$), 5—*AGATTTTCC* ($A_4A_5 \rightarrow T_4T_5$), 6—*AGAAAAACA* (*dlκB*^{P}), 7—*AGAAAATCC* ($T_6 \rightarrow A_6$), 8—cold competition with *AGAAATTCC* probe.

motif) (Figure 6A). Although A_6 of both repressor motifs *AGAAAA*$_6$*TCC* and *AGAAAA*$_6$*ACA* displayed similar opening values, it was significantly higher than that for the activator motif *AGAAAT*$_6$*TCC* (Figure 6B). Opening angle is much larger for A_5, in addition to A_6, of the *AGAAA*$_5$*A*$_6$*ACA* motif (Figure 6B). Since, DNA containing ≥3A–T pairs in a row has displaced N_6 of 'A' and O_4 of 'T' towards the 3′-ends, we measured O_4–N_6 hydrogen bond in different *κB*-motifs (27–29). Interestingly, the two repressor motifs *dlκB*^{P} (*AGAAAAA*$_6$*ACA*) and *zen* (*GGAAAA*$_6$*TCC*), which have ≥4A, displayed increased O_4–N_6 hydrogen bond for A_6 compared to activator motifs, which have T_6 at the corresponding base position (only underlined bases are shown on the X-axis) (Figure 6C). We found that *dlκB*^{P} motif has two abrogated O_4–N_6 hydrogen bonds (A_5 and A_6) in contrast to only one in the *zen* motif (A_6). As a result, the major groove geometry of *dlκB*^{P} motif is more constrained compared to that of the *zen* motif (Figures 5B, C and 6C).

Our analysis of different DNA local parameters revealed dramatic changes in values of the opening angle (Figure 6B). Because of a high opening angle, the A_5–T_5 pair of the *AGAAAAACA* motif is so far stretched that it does not form the O_4–N_6 hydrogen bond, resulting in loss of major groove at the fourth and fifth positions (Figures 3B, D, 6B and C). A–T pairs in an A-tract are weak and have the tendency to form hydrogen bond diagonally (bifurcated hydrogen bond) to stabilize the DNA, as seen in binding of λ434 repressor to its operator motif (13). However, no bifurcated hydrogen bond was seen in the repressor motifs (Figure 5B and C). Formation of bifurcated hydrogen bond in an A-tract

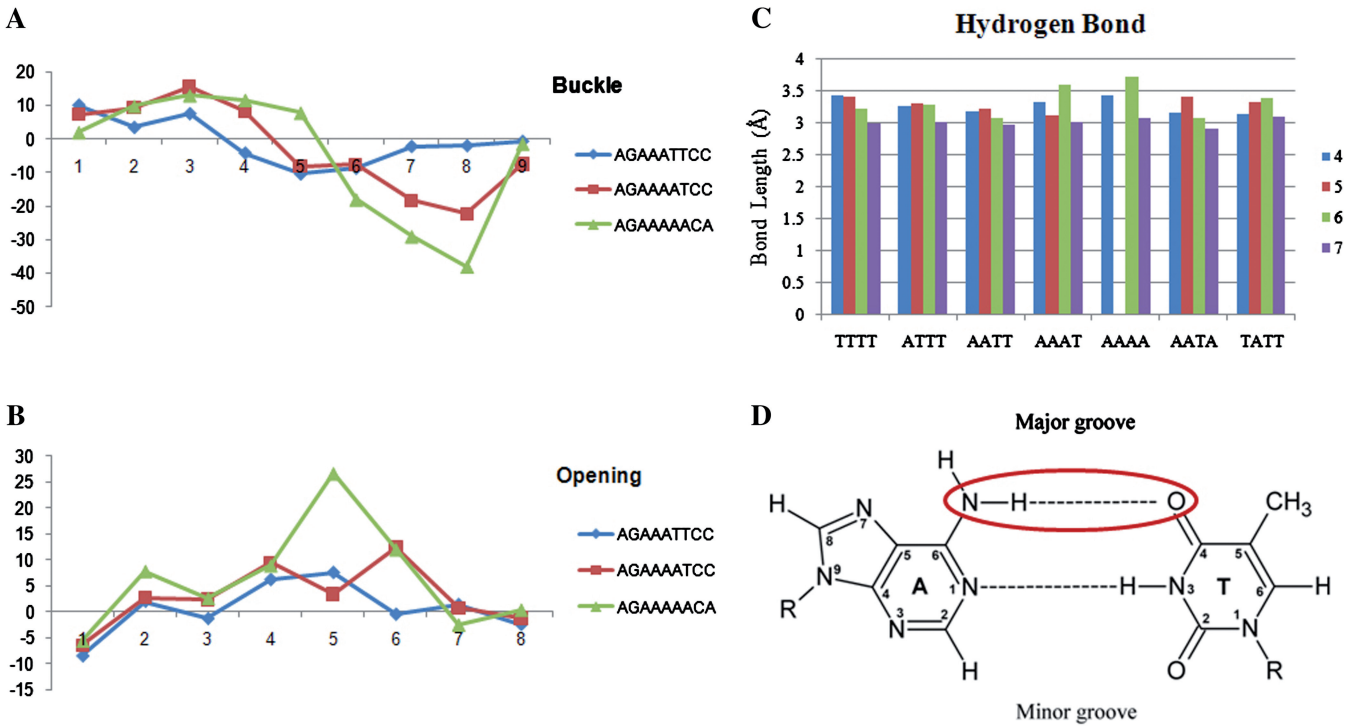

Figure 6. Repressor motifs have weak A-T pairing. (**A** and **B**) Longer A-tracts of repressor motifs (AAAT, AAAAA) show more buckling (5A > 4A > 3A-tract) and large opening in the second half-sites. Large opening angles of A_5 and A_6 correspond to reduced major groove width at these positions (Supplementary Figure S6). Cartoon shows opening angle in nucleotide pair. (**C**) N_6–O_4 hydrogen bond was measured for fourth to seventh nucleotides. It is longer (>3.5 Å) at sixth A in the two repressor motifs indicating weakening of the A_6–T_6 bond in them. Due to extremely large opening angle, A_5 of *AGAAAAACA* does not form N_6–O_4 hydrogen bond which results in acute narrowing of the major groove at A_5 (Figure 3D and Supplementary Figure S6). Also see Table 1 data which highlights stacking properties of A–T pairing in repressor and activator motifs. (**D**) This diagram shows that N_6–O_4 hydrogen bond of the A–T pair faces the major groove.

motif leads to propeller twisting which in turn results in narrowing of the minor groove (27–29).

In our view, absence of bifurcated hydrogen bond in the *AGAAAAACA* motif (which has A-tract) can explain why there are no significant differences in the minor groove of the repressor and the activator motifs (Figure 4A and Supplementary Figure S7). This further suggests that in absence of any significant changes in the minor groove, the unique structure of the repressor motifs is most probably on account of loss of O_4–N_6 hydrogen bonds in the major groove of the second half-site. Activator motifs, in contrast, have all the hydrogen bonds intact (Figure 6C). This finding lends credence to our hypothesis that owing to the presence of A-tract in *κB*-motifs with AAA_6A or AAA_7T sequence, structure of such motifs is more deformable compared to that of AAT_6T type motifs.

Additionally, intramolecular hydrogen bonds are also known to impart specificity. For example, an intramolecular hydrogen bond (cytosine N_4 to a neighbouring phosphate) has been shown to be critical for yeast phenylalanine tRNA function (53,54). Also, A–T and T–A base reversals are more sensitive to major groove changes (55–56). Since O_4–N_6 hydrogen bond is subject to precise stereochemical constraints in an A–T or T–A base pair and hence it is the probable read out of the structural difference that can discern deformable (repressor) and non-deformable (activator) *κB*-motifs.

We surmise that presence of T_6 in the activator motifs breaks the 'A-run', resulting in normal O_4–N_6 hydrogen bond and stable DNA conformation (Figures 5D–F and 6C). Thus, our data suggest that A-tract in a *κB*-motif confers a structural deformability without involving bifurcated hydrogen bond or minor groove narrowing. We propose that absence of bifurcated hydrogen bond or minor groove narrowing is compensated by interaction of Dl with the co-repressor protein.

Identification of co-regulator dependent expression of Dl target genes

Next, we attempted to understand if there is any evolutionary pattern in the sequence composition of *κB*-motifs. In an earlier study, Copley *et al.* (57) reported evolutionary dynamics of *κB*-motifs. There are 4028 *κB*-motifs that are conserved in different *Drosophila* species. We identified 315/4028 motifs as A-tract *κB*-motifs (minimum 4 'A') (Supplementary Table S1). Chromosomal distribution of A-tract *κB*-motifs with respect to 4028 conserved *κB*-motifs did not suggest any chromosomal bias in the distribution of A-tract *κB*-motifs (Figure 7A). Our analysis revealed that 126 of the 315 A-rich *κB*-motifs were highly conserved, and these were carried forward for further analysis (cut-off 0.8) (Figure 7A). A majority of these A-tract *κB*-motifs (86/126) were repressor type (A-tract motifs where A is the sixth nucleotide) while 36/126 were of activator-type (A-tract motifs where sixth nucleotide

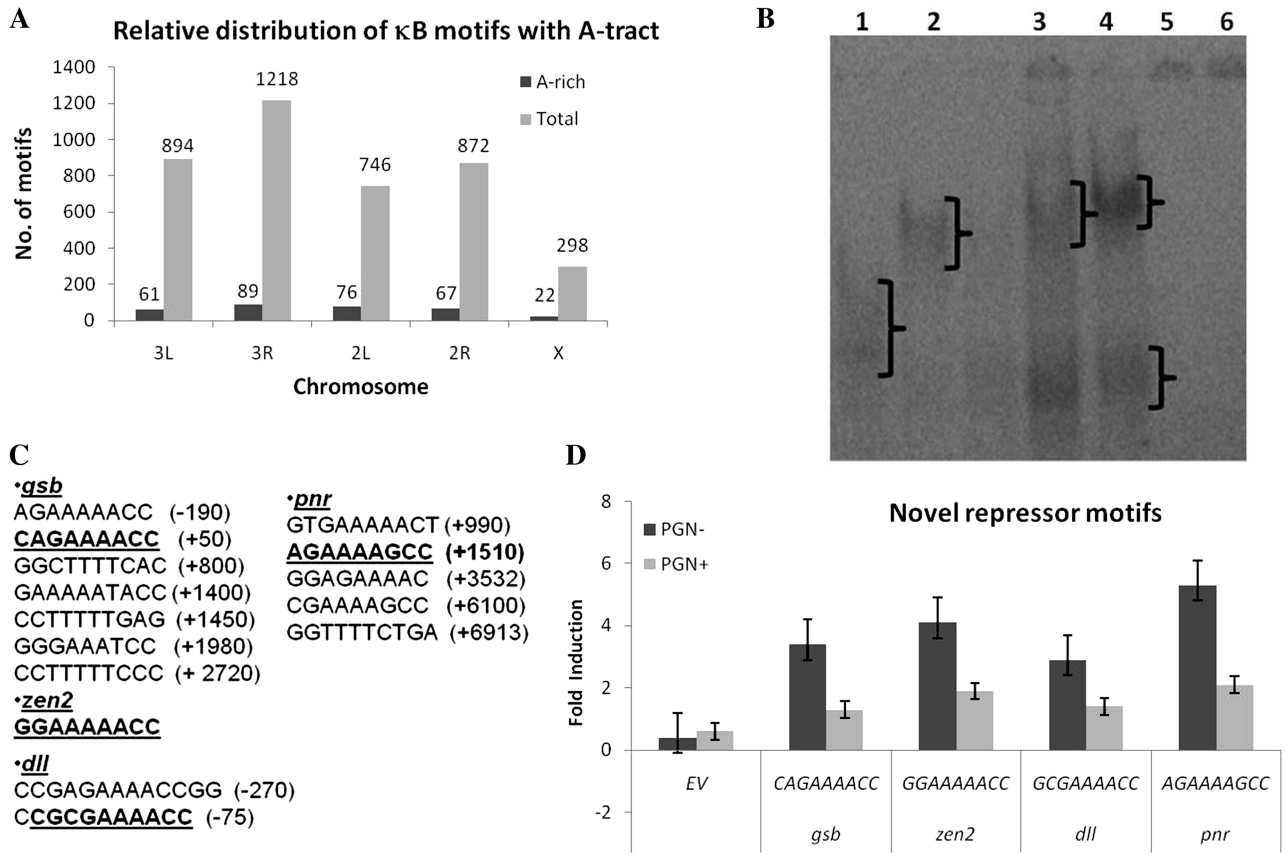

Figure 7. Annotation of repressor *κB*-motifs in *Drosophila* genome. (**A**) Out of 4029 conserved *κB*-motifs, 315 were found to have A-tract (minimum 4A). Their distribution did not suggest any chromosomal bias. (**B**) EMSA was performed with novel A-tract *κB*-motifs, identified in this study, which retarded complexes of different sizes. (Lane 1—*zen2* probe, lane 2—*dll* probe, lane 3—*pnr* probe, lane 4—*gsb* probe, lane 5—mutant probe, lane 6—free probe). Smaller size complexes retarded in lanes 3 and 4 correspond to activator motif $dl\kappa B^{I}$ *(GGGAATTCC)*. The protocol for the competitive EMSA with two different probes (lanes 4 and 5) is explained in reference 39. (**C**) Different *κB*-motifs present in *gsb*, *dll*, *pnr* and *zen2* are indicated along with their position with respect to TATA element. Motifs which retarded Dl-DNA complexes in EMSA in Figure 7B are putatively functional and are shown in bold and are underlined. (**D**) Loss/reduction of luciferase expression upon Dl activation consequent to peptidoglycan (PGN) treatment confirms repressor function of the functional *κB*-motifs of *gsb*, *dll*, *pnr* and *zen2* genes.

was T) (Supplementary Table S2). The remaining four motifs in this list are probably non-functional as they lack the most crucial second position G, which is typically present in all *κB*-motifs characterized till now. In fact, these four motifs completely lack G at any position (Supplementary Table S2). Presence of G in the first half-site and C in the second half-site is a special feature of functional *κB*-motifs (Figure 1).

The list of 86 repressor-type *κB*-motifs included those present in known Dl-repressed genes like *dpp*, *zen* etc, which suggests that other genes featuring in this list might be repressor targets of Dl as well (Supplementary Table S1). Dl-activated and -repressed genes can also be loosely identified by their spatial expression along the dorso-ventral axis. In early embryogenesis, the Dl-repressed genes are expressed in the dorsal ectoderm while Dl-activated genes are expressed on the ventral side. Using a whole-genome approach Stathopoulos *et al.* (46) identified novel Dl target genes that are expressed in the ectoderm. According to our findings, ectodermal genes that are repressed by Dl should have A_6-type *κB*-motifs, and indeed we found that all these genes have $\mathbf{A_6}$ACC as second half-site in their respective *κB*-motifs (Supplementary Table S2).

Next, we examined physical interaction of Dl with the *κB*-motifs of four representative genes namely *gsb*, *pnr*, *dll* and *zen2*. These genes have multiple Dl-binding sites, and here we identified the functional *κB*-motifs in them by EMSA (Figure 7B and C). We found that the functional *κB*-motifs of the four genes retarded bigger DNA-Dl complexes compared to the activator motif *GGGAATTC C*, which was used as a control, indicating the presence of additional proteins in the Dl–DNA complexes (Figure 7C). The functional *κB*-motifs of all the four genes have 'A-tract' with A as the sixth nucleotide (Figure 7C). Furthermore, luciferase assay also proves that Dl-mediated repression of these *κB*-motifs indicating role of co-regulator-mediated repression of target genes by Dl (Figure 7D).

According to the currently accepted model of Dl regulation, the co-regulator-dependent gene regulation by Dl may lead to repression of the target gene expression. Whether gene activation by Dl requires a co-regulator or not, needs to be investigated further but such a possibility

cannot be ruled out. It will be interesting to see if co-regulator dependent gene activation by Dl would also require A-tract in the *κB*-motif for the binding of Dl–co-activator protein complex as is the case with the binding of the Dl–co-repressor complexes.

DISCUSSION

The mechanisms that control the precisely regulated switch from gene repression to gene activation represent a central question in transcriptional regulation. One feature of this transcriptional reorganization is the cross-talk with co-factor. It is understood that activating stimuli induce recruitment of TFs along with their co-activators to target gene promoters while a repressor signal leads to the assembly of co-repressors at the gene promoter. In general, co-repressors and co-activators most often act in a gene specific manner. Although it is known that gene activation or suppression by bi-functional TFs is context dependent, it is not clear what factors determine that such TFs interact with a co-repressor at one gene and with a co-activator at another.

Co-regulator specificity in Dl binding

The current paradigm is that interaction of a TF with its co-factor is dependent only on protein–protein interaction (58). Here we report that Dl interaction with its co-regulator is also decided by the sequence of the Dl-binding DNA motif. An analysis of activator and repressor Dl motifs suggested that activator or repressor function of Dl might be encoded in the sequence of the *κB*-DNA motif. Interestingly, Dl–co-regulator interaction was seen with the repressor motifs but not with the activator motifs, implying that repressor function of Dl is probably co-regulator dependent (Figures 3 and 8). Previous reports have also suggested that transcriptional repression by Dl is co-regulator dependent (51,39). Our study suggests that nucleotide at the sixth position is an important determinant of not only Dl binding but also of its function as a transcriptional activator or repressor (Figures 2 and 3). Importance of sixth T of a *κB*-motif in binding with Rel proteins has been implicated in at least two previous studies involving 17 mammalian *κB*-motifs using biochemical approaches (49,59). The only exception to this rule is the *κB*-binding motif of the IL2 receptor, which has C in place of T at the sixth position. Although Muroi *et al.* (49) showed the importance of sequence composition in binding of Rel proteins, the mechanism by which the *κB* sequence imparts specificity has remained unclear.

Our study reveals that T_6 motifs characterize co-regulator-independent transcriptional regulation by Dl, whereas 'A' at the sixth position in a *κB*-motif signifies a co-regulator-dependent regulation by Dl. This specificity is on account of Dl–co-repressor complex binding on an 'A-tract'*κB*-motif (A-tract is due to presence of 'A' at the sixth position, as a result there are four continuous 'As' in the repressor *κB*-motifs) (Figure 2D). On the other hand, activator motifs (T_6-*κB*-motif) have 'T' at the sixth position which breaks the run of continuous 'A' residues on the same strand. This is the distinguishing feature of the two motifs.

Role of *κB*-DNA sequence in determining Dl-co-factor specificity

A tract of four to six consecutive (dA): (dT) residues imparts typical narrow and straight minor groove geometry to DNA (28,29). However A-tract *κB*-motifs have reduced major groove at certain points which, in our opinion, is due to partial structural collapse triggered by loss of O_4–N_6 hydrogen bond. The A-tract has an inherent property to bend but a single nucleotide insertion that interrupts the run of A also breaks this bending (60). Thus an *AAAA*/*AAAT* motif is more deformable than *AATT* or *AATA* motif. Furthermore, co-repressor binding is favoured by DNA sequences for which conformational constraints have low energy requirements. Accordingly co-repressor binding is more favoured on *AAA* motif than it is on *ATA*, as seen in the case of interaction between bacteriophage λ434 repressor with its operator (61). It is for this reason that repressor motifs that have *AAAA*/*AAAT* core can accommodate more structural distortions in order to facilitate binding of Dl in complex with different co-regulators. A similar mechanism has been proposed for differential binding affinity of λ434 operator in complex with different repressor and Cro proteins (62). While the typical geometry of the λ434 operator is stabilized by a bifurcated hydrogen bond, such a bond is not seen in the *AGAAAAACA* motif. Interestingly the unique geometry of *AGAAAAACA* motif is due to kink at the hinge base (A_5), and based on our biophysical and biochemical data we propose that this geometry is important for binding of Dl–AP1 complex to this motif (Figures 5G and 6A–D).

Our study provides structural insights into the nucleotide preferences in co-regulator-dependent and -independent transcriptional regulation by Dl. This serves to highlight the unique ways by which DNA can exert specificity not only on DNA–protein interactions but also on interactions between regulator and co-regulator. There is evidence also that alterations in the overall structure of DNA-binding domains can influence the DNA sequence preferences in numerous ways (63,64). Previous work has shown that co-operative binding of ATF-2/c-Jun and IRF3 depend on inherent asymmetry of the ATF-2/c-Jun-binding site, so that mutation of the latter to a consensus AP1 recognition element resulted in loss of interaction (65). In another study, interaction of Fos-Jun heterodimer with NFAT on its binding site on the IL-2 enhancer was shown to be dependent on the co-operativity 'at some level of assembly' between the three proteins and the 'DNA backbone' (66). It is believed that assembly of proteins on DNA is a result of co-operativity between DNA and the binding proteins and that co-operative binding can arise through nucleotide sequence-guided structural changes in the DNA which allow formation of complementary DNA conformations for the adjacently bound TFs (67–69).

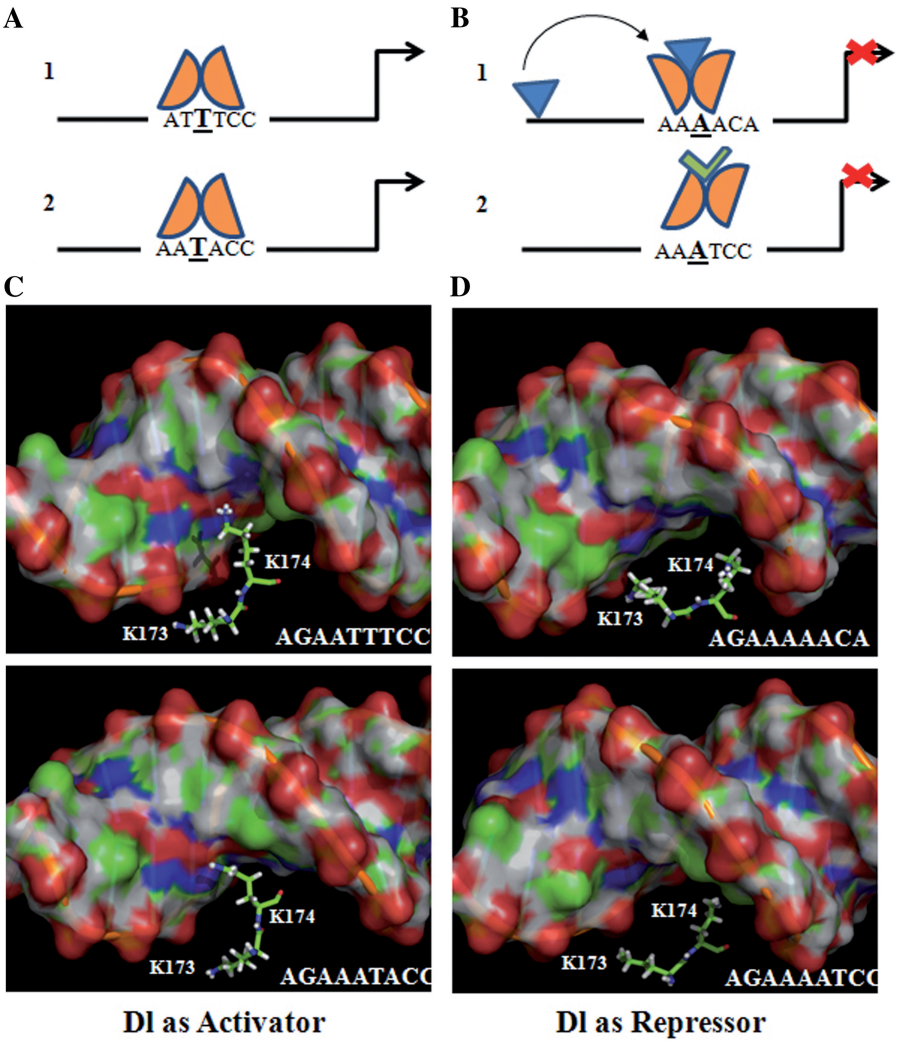

Figure 8. Putative model of transcriptional activation and repression by Dl. (**A**) Different activator *κB*-motifs (sixth T is common) retard similar size Dl-complexes suggesting that gene activation by Dl is probably independent of co-regulators. (**B**) Gene repression by Dl is context-based and co-repressor dependent as different repressor motifs (sixth A is common) retard Dl-complexes of varying sizes indicating presence of proteins of different sizes (also see Figures 2 and 5). A-tract core in the repressor motifs has intrinsic property to bend. This intrinsic deformability in repressor motifs is sequence specific which allows specific recruitment of Dl–co-repressors complexes in gene specific manner as shown in the schematic. Dl interaction with the co-regulator could be either *cis* (B1, upper panel) or *trans* (B2, lower panel). (**C** and **D**) For successful binding, the DNA–protein interface should match molecular surfaces which includes protein contacts to not only base pairs in the major groove but also to sugar–phosphate backbone. We have shown that activator and repressor motifs have different major groove conformations. As a result, Dl interaction with the two motifs is different (e.g. interaction of K173 and K174 of Dl with the two activator motifs is very similar). On the other hand, binding of the same two residues with the two repressor motifs is very different. Hence, we opine that 'A-tract *κB*' motifs have deformable structure to facilitate context-based interaction of the same Dl with different co-repressors in gene-specific manner as revealed by differences in their DNA geometry.

Mammalian *κB*-motifs have comparable DNA shape

Our analysis of major groove geometries of *κB*-motifs for which crystal structures are available indicated a rather similar DNA geometry. All these *κB*-motifs have their major groove maxima at second and eighth positions and minima at the fifth position that gives the stretched eagle geometry (Supplementary Figure S10). However, none of these motifs can be classified as A-tract containing and thus a crystal structure of Rel monomers bound to an 'A-tract' κB-motif is still not known. In a typical *κB*-motif, nucleotides at the termini are involved in DNA–protein interactions and are more conserved than core nucleotides GGRNWTTCC (underlined bases indicate core nucleotides) (Figure 1A). Surprisingly, DNA geometry of these terminal nucleotides is not so conserved (Supplementary Figure S10). On the contrary major groove structures of the core nucleotides (positions 4–6), where sequence conservation is the least, is relatively high (Figure 1 and Supplementary Figure S10). Maximum conservation of major groove width was observed for the hinge base (Supplementary Figure S10). This suggests that these *κB*-motifs, in spite of sequence differences, have common structural paradigm (Supplementary Figure S10). Conserved geometry of major grooves at positions 4–6 was unexpected (Supplementary Figure S6). Nitrogenous bases at these positions do not form hydrogen bonds with Rel proteins (direct contact)

although they are involved in electrostatic interactions with different Rel monomers and hence conserved geometry phosphate backbone at these positions was intriguing. In our view, conserved geometry of core nucleotides is essential as it provides unique framework to *κB*-DNA. We have shown that mutations that affect DNA geometry in this region are associated with strong phenotypes (Figure 2 and Supplementary Figure S6). If electrostatic interactions involving phosphate backbone of protein non-contacted bases were not important then $T_6 \rightarrow A_6$ mutation should not have affected repressor/activator function of Dl or its interaction with co-repressor (Figures 2–5 and Supplementary Figure S6). These results suggest that phosphate backbone geometry is sequence dependent and guides Rel–DNA binding. Thus, we have revealed distinct roles played by protein contacted and protein non-contacted nucleotides in a *κB*-motif.

κB-DNA geometry explains permissiveness in Rel binding

DNA-binding RHD is highly conserved, yet different Rel proteins bind *κB*-motifs of their respective target genes without overlap (70). This raises a question that how different Rel proteins identify their binding motifs specifically. We suggest that specificity in Rel–DNA interaction is regulated by protein non-contacted 'core' as well as protein contacted 'terminal' nucleotides in respective Rel binding motifs. A similar example of DNA–protein interaction is seen in the HPV E2 system in which the 'A-tract' core, which is also not contacted by the protein, ensures specificity of interaction (71). The importance of the structure of protein non-contacted nucleotides in DNA–protein interaction is also highlighted by the overwound configuration of central nucleotides of the phage λ434 operator (13).

Although core nucleotides of *κB*-DNA are not contacted by Rel proteins, they play a critical role by imparting the correct conformation to *κB*-DNA, and hence they display a conserved geometry. However, it is the major groove variations at the termini that make each *κB*-motif structurally different from others (Supplementary Figure S10). These unique structures of *κB*-motifs are compatible with one particular homo/hetero-dimer of Rel proteins and not others, which can explain the specificity with which different Rel proteins identify and bind their cognate *κB*-motifs. e.g. base-specific contacts within the p50 and p65 homodimers and p50–p65 hetero-dimer suggest that the two RHDs contained in each dimer relate to each other differently in different structures (Supplementary Figure S10) (72–75). Hence, only Rel isoforms which have the chemically complementary DNA geometry will bind that specific *κB*-motif and not others. We propose that structural variations at the terminal positions may explain how different Rel proteins selectively interact with target gene promoters. Thus, it is the structural compatibility which is more important than the sequence of nucleotides in a *κB*-motif which in turn can explain the evolution of sequence variation in a *κB*-motif (sequence permissiveness).

Transcriptional property of Dl is encoded in its binding motif sequence

Dl, like its mammalian homologues p65, c-Rel and RelB, possesses transcriptional activation domain and hence is classified as transcriptional activator (36,37,75). However, Dl has been shown to activate as well as repress transcription of different genes along the dorso-ventral axis. There are at least two models to explain the ability of Dl to activate or repress target genes. According to one model, Dl by default is an activator and works as a repressor only in specific promoter contexts (51,76,77). However, another model suggests that Dl recognizes two classes of sites which have different allosteric effects on the protein and that can result in either transcriptional activation or repression, e.g. Dl-binding sites in *twi* (Dl as activator) is less symmetrical compared to that in *zen* (Dl as repressor) (78). Even so, it has not been established whether specific sequence variations in the Dl-binding motifs lead to differential regulation of Dl target genes.

The present study reveals the sequence-specific allosteric changes in *κB*-motifs which regulate binding of Dl to precise *κB*-motifs with complementary structure and thus confer specificity not only to Dl–DNA interaction but also to Dl–co-regulator interaction. This might have led to the evolution of distinct sequence heterogeneity in Dl-binding motifs *viz*. 'A-rich' *κB*-motifs in genes repressed by Dl and 'A or T-rich' *κB*-motifs in genes activated by it, i.e. Dl-binding motifs exist in two forms (i) activator, and (ii) repressor conformation (Figure 8). This is also true for another Rel protein, human p65 which is a known transcriptional activator but has been shown to bind to human P-sequence '*GAAAATTTCC*' of IL-4 gene leading to transcriptional repression of IL-4 (A-tract is underlined). Interestingly, the P-sequence motif has an A-tract as well. This suggests that A-tract in a *κB*-motif is required for transcriptional repression by activator human Rel protein p65 as well. Whether co-regulator is involved in transcriptional repression of IL-4, by p65 homodimer or not, is currently not known and needs further investigation.

Further, we have used this structural information to identify and characterize the functional Dl-binding motifs in the *Drosophila* genome and classify them as activator or repressor motifs. Conventionally whole genome and tiling arrays studies have been employed to identify Dl target genes. A major limitation of these approaches is that they do not give information whether the target gene is activated or repressed by Dl. Our molecular dynamics approach circumvents this limitation by predicting the structure of the *κB*-motif as activator/repressor conformations. This can potentially aid in 'functional annotation' of Dorsal gene regulatory network.

Our study, importantly, suggests that specificity in Dl–DNA binding may also derive from interactions involving double helical backbone. Backbone contacts impart specificity in DNA–protein interactions through the positioning of protein recognition elements in orientations that allow them to make other more specific contacts. This is particularly important as sequence dependent deformability of DNA has been observed in

DNA–protein complexes (79,80). Hence, while a typical *κB*-motif is recognized by Dl alone the A-tract *κB*-motifs, due to their unique shapes, are not. As a result such motifs are recognized by Dl only when bound to its co-regulator. It is also possible that sequence-dependent DNA structure may also contribute to co-operative binding of Dl with its co-regulators. This suggests that sequence-dependent DNA structure may be critical for binding of individual factors (e.g. co-operative binding of AP1 and Dl complex on the Dl promoter), but also in the assembly of Dl–AP1 multi-protein complex (39). Further structural studies would be needed to evaluate the proposals made here. The major limiting factor at this stage is the lack of crystal structures of *Drosophila* Rel proteins. Another limitation is the fundamental difficulty in accommodating small movements of amino acid side-chains, which are likely to occur in Dl interaction with its binding motif. Additionally, role of hydrophobic stacking interactions in the recognition process needs to be understood. Stacking interactions are somewhat sequence dependent, and it is not clear at present how the intercalation of planar amino-acid side chains can be used in a DNA–protein recognition system.

Significance of stacking interactions can also be envisaged from the melting temperatures of different *κB*-motifs as shown in Table 1. The T_m was calculated online using T_m Predictor software (http://www.scfbio-iitd.res.in/chemgenome/Tm_predictor.jsp). This programme explicitly accounts for disruption in stacking interactions, breakage of hydrogen bonds apart from other physico-chemical parameters to predict T_m of a given DNA sequence (81). It is evident that the repressor motifs have lesser T_m compared to corresponding activator motif e.g. $dl\kappa B^I$ motif (T_m = 49.29°C) is an activator while $dl\kappa B^P$ motif (T_m = 38.68°C) is a repressor (Figures 2, 3 and Supplementary Figure S6). This T_m difference is significant and can potentially explain the kink in the major groove of the $dl\kappa B^P$ motif (Figures 3D, E, 5B, 6 and 8). It is interesting to note that although both activator and repressor motifs have equal number of A or T nucleotides still T_6-motif has higher T_m compared to A_6 motif which points to the role of stacking interactions. The same pattern is also seen in the repressor *zen* motif (*GGGAAAACC*, T_m = 49.29°C) and its T_6 mutant (*GGGAATACC*, T_m = 50.11°C), which acts as activator. In all these cases the difference in T_m of the respective activator–repressor pair is <1°C (Table 1 and Figure 2A–D). It is not clear whether or not small changes in T_m will have significant impact on DNA–protein interactions, but the present analysis suggests that motifs with low T_m will be more amenable to structural perturbations as compared to motifs with high T_m. We have shown that *AGAAAATCC* motif, where sixth nucleotide is A, facilitates binding of the co-regulator, but interaction of Dl with its co-regulator is lost upon $A_6 \rightarrow T_6$ mutation (Figures 2, 5 and Supplementary Figure S6). These analyses suggest that *κB*-motifs with lower T_m may facilitate Dl–co-regulator interaction probably because there are less stacking interactions in such motifs which make them more amenable to structural perturbation. In other words A-tract motifs are more deformable which might be required for Dl interaction with co-activator/co-repressor proteins (Figure 8).

Table 1. Comparison of melting temperatures (T_m) of activator repressor combinations of different *κB*-motifs

Motif	Sequence	T_m (°C)
$dlkB^I$	*GGGAA**T**TCC* (Act)	49.29
$dlkB^P$	*AGAAAA**A**ACA* (Rep)	38.68
	*AGAAAAA**A**TCC* (Rep)	41.94
	*AGAAA**T**ACC* (Act)	42.76
zen	*GGGAAA**A**ACC* (Rep)	49.29
	*GGGAA**T**ACC* (Act)	50.11
phm	*GGGAT**T**ACC* (Act)	50.11
	*GGGAA**A**ACC* (Rep)	49.29

(Act—Activator; Rep—Repressor).
It is evident that activator motifs have higher T_m compared to corresponding repressor motifs. T_m difference is highest for $dl\kappa B^P$ motif and its corresponding activator mutant and their backbone structures are drastically different (Supplementary Figure S6). This is indicative of the correlation between T_m and structural flexibility of the DNA sequence. This might be an important determinant for interaction of TFs (in combination with or without co-regulators) with cognate DNA-binding motifs.
This is also to be noted that this comparison holds true for the repressor–activator combination of the same *κB*-motif but not for two different motifs. e.g. T_m of wild-type *zen* motif can be compared with *zen*-mutant and similarly wild-type *phm* motif can be compared with *phm* mutant. Nucleotide at the sixth position is indicated in bold while, mutated nucleotide at this position is also underlined.

In recent years, information on the role of nucleosome in gene transcription has begun to emerge. It has been shown that nucleosomes frequently assume specific positions on DNA, which is called as nucleosome positioning (82,83). The signals on DNA for the nucleosome packaging code reside in the structural properties of DNA base pair combinations, indicating the role of conformation of particular DNA sequences in deciding the code (84,85), e.g. positioning of nucleosomes was shown to operate by excluding nucleosomes from 'A-tract' DNA. A-tracts take a context-independent structure, distinct from canonical B-DNA, due to their ability to switch in a cooperative manner (86–88). Because of this A-tracts become a conformationally rigid DNA motifs that constrain B-DNA regions bordering them (29). Furthermore, short A-tracts stabilize the deformation required for histone facing nucleosomal DNA as a result of which DNA bending, deformability and other shape readouts become important in nucleosome positioning (89).

Currently it is not known whether Dl binding to its cognate motif requires a nucleosome-free region or nucleosome-bound DNA. To study the role of nucleosomes in transcriptional regulation by Dl would require the knowledge of nucleosome positions in target genes. Such a study would also uncover the role of sequence composition on nucleosome positioning in relation to Dl binding, if any. We have recently shown that swapping of a repressor motif with an activator motif or vice versa does not affect the transcriptional property of the respective motifs in autoregulation of *dl* gene by Dl (39). Thus one can speculate that Dl binding to its cognate activator

or repressor motifs in the *dl* gene, at least, might be independent of nucleosome position.

Current study suggests that the geometry of κB-motif may determine the specificity with which Dl identifies functional regulatory sites. In addition to providing molecular understanding of DNA recognition by Dl, our data also provide a protocol for identifying functional regulatory κB-motifs. This study also reveals the structural basis of the sequence code which can discriminate a functional κB-motif as an enhancer or as a repressor and also highlights the allosteric changes that κB-motifs induce on DNA–protein interaction. The significant aspect of this study is the elucidation of deformable nature of κB-motifs where a co-regulator is involved. This structural deformability is on account of increase in or complete loss of the major groove facing hydrogen bond in Rel-binding DNA motifs. In conclusion, we show that Rel proteins have distinct κB-sequence preferences according to their biological functions. We surmise that evolution of sequence diversity in κB-motifs is not random, and that it has occurred under selection pressure from Rel-co-regulator network in a gene-specific manner.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are thankful to Thomas Bishop, Tulane University and B. Jayaram of Indian Institute of Technology, Delhi, for their comments and suggestions.

FUNDING

Centre of Excellence grant from Department of Biotechnology, Government of India (to J.N.); Council for Scientific and Industrial Research, Government of India Research Fellowship (to N.M.). Funding for open access charge: Centre for DNA Fingerprinting and Diagnostics.

Conflict of interest statement. None declared.

REFERENCES

1. Aggarwal,A.K., Rodgers,D.W., Drottar,M., Ptashne,M. and Harrison,S.C. (1988) Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science*, **242**, 899–907.
2. Mandel-Gutfreund,Y., Schueler,O. and Margalit,H. (1995) Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. *J. Mol. Biol.*, **253**, 370–382.
3. Pabo,C.O. and Nekludova,L. (2000) Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? *J. Mol. Biol.*, **301**, 597–624.
4. Kono,H. and Sarai,A. (1999) Structure-based prediction of DNA target sites by regulatory proteins. *Proteins*, **35**, 114–131.
5. Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
6. Jones,S., van Heyningen,P., Berman,H.M. and Thornton,J.M. (1999) Protein-DNA interactions: a structural analysis. *J. Mol. Biol.*, **287**, 877–896.
7. Saenger,W. (1983) Water and nucleic acids. In *Principles of Nucleic Acid Structure*, Vol. 17. Springer, New York, pp. 368–84.
8. Shakked,Z. and Rabinovich,D. (1986) The effect of the base sequence on the fine structure of the DNA double helix. *Prog. Biophys. Mol. Biol.*, **47**, 159–195.
9. Luscombe,N.M., Austin,S.E., Berman,H.M. and Thornton,J.M. (2000) An overview of the structures of protein-DNA complexes. *Genome Biol.*, **1**, 1–37.
10. Luscombe,N.M. and Thornton,J.M. (2002) Protein–DNA interactions: amino acid conservation and the effects of mutations on binding specificity. *J. Mol. Biol.*, **320**, 991–1009.
11. Otwinowski,Z., Schevitz,R.W., Zhang,R.G., Lawson,C.L., Joachimiak,A., Marmorstein,R.Q., Luisi,B.F. and Sigler,P.B. (1988) Crystal structure of trp repressor/operator complex at atomic resolution. *Nature*, **335**, 321–329.
12. Shakked,Z., Guerstein,G., Frollow,F., Rabinivich,D., Joachimiak,A. and Sigler,P.B. (1994) Determinants of repressor-operator recognition from the structure of the trp operator binding site. *Nature*, **368**, 469–473.
13. Koudelka,G.B., Harrison,S.C. and Ptashne,M. (1987) Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro. *Nature*, **326**, 886–888.
14. Schultz,S.C., Shields,G.C. and Steitz,T.A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science*, **253**, 1001–1007.
15. Nickolov,D.B., Chen,H., Halay,E.D., Hoffman,A., Roeder,R.G. and Burley,S.K. (1996) Crystal structure of a human TATA box–binding protein/TATA element complex. *Proc. Natl Acad. Sci. USA*, **93**, 4862–4867.
16. Rice,P.A., Yang,S.W., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHFDNA complex: a protein-induced DNA U-turn. *Cell*, **87**, 1295–1306.
17. Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.
18. Kim,Y., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
19. Flatters,D. and Lavery,R. (1998) Sequencedependent dynamics of TATA-Box binding sites. *Biophys. J.*, **75**, 372–381.
20. Travers,A. (ed.), (1993) *DNA-Protein* Interactions, 1st edn. Chapman & Hall, London.
21. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of Specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
22. Jayaram,B. and Jain,T. (2004) The role of water in protein-DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 343–361.
23. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
24. Leslie,A.G., Arnott,S., Chandrasekaran,R. and Ratliff,R.L. (1980) Polymorphism of DNA double helices. *J. Mol. Biol.*, **143**, 49–72.
25. Shakked,Z., Guerstein-Guzikevich,G., Eisenstein,M., Frolow,F. and Rabinovich,D. (1989) The conformation of the DNA double helix in the crystal is dependent on its environment. *Nature*, **342**, 456–460.
26. Ng,H.L., Kopka,M.L. and Dickerson,R.E. (2000) The structure of a stable intermediate in the A $\leftrightarrow$ B DNA helix transition. *Proc. Natl Acad. Sci. USA*, **97**, 2035–2039.
27. Nelson,H.C., Finch,J.T., Luisi,B.F. and Klug,A. (1987) The structure of an oligo(dA)•oligo(dT) tract and its biological implications. *Nature*, **330**, 221–226.
28. Hizver,J., Rozenberg,H., Frolow,F., Rabinovich,D. and Shakked,Z. (2001) DNAbending by an adenine-thymine tract and its role in gene regulation. *Proc. Natl Acad. Sci. USA*, **98**, 8490–8495.
29. Haran,T.E. and Mohanty,U. (2009) The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.*, **42**, 41–81.
30. Jayaram,B., Sharp,K. and Honig,B. (1989) The electrostatic potential of B-DNA. *Biopolymers*, **28**, 975–993.

31. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
32. Badis,G., Berger,M.F., Philippakis,A.A., Talukder,S., Gehrke,A.R., Jaeger,S.A., Chan,E.T., Metzler,G., Vedenko,A., Chen,X. *et al.* (2009) Diversity and complexity in DNA recognition by transcription factors. *Science*, **324**, 1720–1723.
33. Joshi,R., Passner,J.M., Rohs,R., Jain,R., Sosinsky,A., Crickmore,M.A., Jacob,V., Aggarwal,A.K., Honig,B. and Mann,R.S. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell*, **131**, 530–543.
34. Margolin,J.F., Friedman,J.R., Meyer,W.K., Vissing,H., Thiesen,H.J. and Rauscher,F.J. 3rd (1994) Kruppel-associated boxes are potent transcriptional repression domains. *Proc. Natl Acad. Sci. USA*, **91**, 4509–4513.
35. Mrinal,N. and Nagaraju,J. (2008) Intron loss is associated with gain of function in the evolution of the gloverin family of antibacterial genes in Bombyx mori. *J. Biol. Chem.*, **283**, 23376–23387.
36. Ip,Y.T., Kraut,R., Levine,M. and Rushlow,C.A. (1991) The dorsal morphogen is a sequence-specific DNA-binding protein that interacts with a long-range repression element in Drosophila. *Cell*, **64**, 439–446.
37. Rushlow,C. and Warrior,R. (1992) The rel family of proteins. *Bioessays*, **14**, 89–95.
38. Leung,T.H., Hoffmann,A. and Baltimore,D. (2004) One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. *Cell*, **118**, 453–464.
39. Mrinal,N. and Nagaraju,J. (2010) Dynamic repositioning of dorsal to two different kappaB motifs controls its autoregulation during immune response in Drosophila. *J. Biol. Chem.*, **285**, 24206–24216.
40. Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
41. Huang,D.B., Chen,Y.Q., Ruetsche,M., Phelps,C.B. and Ghosh,G. (2001) X-ray crystal structure of proto-oncogene product c-Rel bound to the CD28 response element of IL-2. *Structure*, **9**, 669–678.
42. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
43. Phillips,J.C., Braun,R., Wang,W., Gumbart,J., Tajkhorshid,E., Villa,E., Chipot,C., Skeel,R.D., Kale,L. and Schulten,K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
44. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38, 27–38.
45. Senger,K., Armstrong,G.W., Rowell,W.J., Kwan,J.M., Markstein,M. and Levine,M. (2004) Immunity regulatory DNAs share common organizational features in Drosophila. *Mol. Cell*, **13**, 19–32.
46. Stathopoulos,A., Van Drenth,M., Erives,A., Markstein,M. and Levine,M. (2002) Whole-genome analysis of dorsal-ventral patterning in the Drosophila embryo. *Cell*, **111**, 687–701.
47. Papatsenko,D. and Levine,M. (2005) Quantitative analysis of binding motifs mediating diverse spatial readouts of the Dorsal gradient in the Drosophila embryo. *Proc. Natl Acad. Sci. USA*, **102**, 4966–4971.
48. Biemar,F., Nix,D.A., Piel,J., Peterson,B., Ronshaugen,M., Sementchenko,V., Bell,I., Manak,J.R. and Levine,M.S. (2006) Comprehensive identification of Drosophila dorsal-ventral patterning genes using a whole-genome tiling array. *Proc. Natl Acad. Sci. USA*, **103**, 12763–12768.
49. Muroi,M., Muroi,Y., Yamamoto,K. and Suzuki,T. (1993) Influence of 3′ half-site sequence of NF-kappa B motifs on the binding of lipopolysaccharide-activatable macrophage NF-kappa B proteins. *J. Biol. Chem.*, **268**, 19534–19539.
50. Zhang,H. and Levine,M. (1999) Groucho and dCtBP mediate separate pathways of transcriptional repression in the Drosophila embryo. *Proc. Natl Acad. Sci. USA*, **96**, 535–540.
51. Kirov,N., Zhelnin,L., Shah,J. and Rushlow,C. (1993) Conversion of a silencer into an enhancer: evidence for a co-repressor in dorsal-mediated repression in Drosophila. *EMBO J.*, **12**, 3193–3199.
52. Seeman,N.C., Rosenberg,J.M. and Rich,A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.
53. Quigley,G.J., Seeman,N.C., Wang,A.H., Suddath,F.L. and Rich,A. (1975) Yeast phenylalanine transfer RNA: atomic coordinates and torsion angles. *Nucleic Acids Res.*, **2**, 2329–2341.
54. Ladner,J.E., Jack,A., Robertus,J.D., Brown,R.S., Rhodes,D., Clark,B.F. and Klug,A. (1975) Structure of yeast phenylalanine transfer RNA at 2.5 A resolution. *Proc. Natl Acad. Sci. USA*, **72**, 4414–4418.
55. Goodsell,D.S., Kaczor-Grzeskowiak,M. and Dickerson,R.E. (1994) The crystal structure of C-C-A-T-T-A-A-T-G-G. Implications for bending of B-DNA at T-A steps. *J. Mol. Biol.*, **239**, 79–96.
56. Gorin,A.A., Zhurkin,V.B. and Olson,W.K. (1995) B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.*, **247**, 34–48.
57. Copley,R.R., Totrov,M., Linnell,J., Field,S., Ragoussis,J. and Udalova,I.A. (2007) Functional conservation of Rel binding sites in drosophilid genomes. *Genome Res.*, **17**, 1327–1335.
58. Diamond,M.I., Miner,J.N., Yoshinaga,S.K. and Yamamoto,K.R. (1990) Transcription factor interactions: selectors of positive or negative regulation from a single DNA element. *Science*, **249**, 1266–1272.
59. Zabel,U., Schreck,R. and Baeuerle,P.A. (1991) DNA binding of purified transcription factor NF-kappa B. Affinity, specificity, Zn2+ dependence, and differential half-site recognition. *J. Biol. Chem.*, **266**, 252–260.
60. Mack,D.R., Chiu,T.K. and Dickerson,R.E. (2001) Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: a comparative analysis of T-A and A-T steps within A-tracts. *J. Mol. Biol.*, **312**, 1037–1049.
61. Koudelka,G.B., Harbury,P., Harrison,S.C. and Ptashne,M. (1988) DNA twisting and the affinity of bacteriophage 434 operator for bacteriophage 434 repressor. *Proc. Natl Acad. Sci. USA*, **85**, 4633–4637.
62. Rodgers,D.W. and Harrison,S.C. (1993) The complex between phage 434 repressor DNA-binding domain and operator site OR3: structural differences between consensus and non-consensus half-sites. *Structure*, **1**, 227–240.
63. Miller,J.C. and Pabo,C.O. (2001) Rearrangement of side-chains in a Zif268 mutant highlights the complexities of zinc finger-DNA recognition. *J. Mol. Biol.*, **313**, 309–315.
64. Wolfe,S.A., Grant,R.A., Elrod-Erickson,M. and Pabo,C.O. (2001) Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure*, **9**, 717–723.
65. Panne,D., Maniatis,T. and Harrison,S.C. (2004) Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. *EMBO J.*, **23**, 4384–4393.
66. Chen,L., Glover,J.N., Hogan,P.G., Rao,A. and Harrison,S.C. (1998) Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature*, **392**, 42–48.
67. Escalante,C.R., Brass,A.L., Pongubala,J.M., Shatova,E., Shen,L., Singh,H. and Aggarwal,A.K. (2002) Crystal structure of PU.1/IRF-4/DNA ternary complex. *Mol. Cell*, **10**, 1097–1105.
68. Klemm,J.D. and Pabo,C.O. (1996) Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes Dev.*, **10**, 27–36.
69. Panne,D., Maniatis,T. and Harrison,S.C. (2007) An atomic model of the interferon-beta enhanceosome. *Cell*, **129**, 1111–1123.
70. Lenardo,M.J. and Baltimore,D. (1989) NF-kappa B: a pleiotropic mediator of inducible and tissue-specific gene control. *Cell*, **58**, 227–229.
71. Hizver,J., Rozenberg,H., Frolow,F., Rabinovich,D. and Shakked,Z. (2001) DNA bending by an adenine–thymine tract and its role in gene regulation. *Proc. Natl Acad. Sci. USA*, **98**, 8490–8495.
72. Ghosh,G., van Duyne,G., Ghosh,S. and Sigler,P.B. (1995) Structure of NF-kappa B p50 homodimer bound to a kappa B site. *Nature*, **373**, 303–310.
73. Chen,F.E., Huang,D.B., Chen,Y.Q. and Ghosh,G. (1998) Crystal structure of p50/p65 heterodimer of transcription factor NF-kappaB bound to DNA. *Nature*, **391**, 410–413.

74. Chen,Y.Q., Ghosh,S. and Ghosh,G. (1998) A novel DNA recognition mode by the NF-kappa B p65 homodimer. *Nat. Struct. Biol.*, **5**, 67–73.
75. Cramer,P., Larson,C.J., Verdine,G.L. and Muller,C.W. (1997) Crystal structure of the human NFkB p52 homodimer –DNA complex at 2.1Å resolution. *EMBO J.*, **16**, 7078–7090.
76. Jiang,J., Rushlow,C.A., Zhou,Q., Small,S. and Levine,M. (1992) Individual dorsal morphogen binding sites mediate activation and repression in the Drosophila embryo. *EMBO J.*, **11**, 3147–3154.
77. Pan,D. and Courey,A.J. (1992) The same dorsal binding site mediates both activation and repression in a context-dependent manner. *EMBO J.*, **11**, 1837–1842.
78. Thisse,C., Perrin-Schmitt,F., Stoetzel,C. and Thisse,B. (1991) Sequence-specific transactivation of the Drosophila twist gene by the dorsal gene product. *Cell*, **65**, 1191–1201.
79. Olson,W.K., Gorin,A.A., Lu,X.J., Hock,L.M. and Zhurkin,V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163–11168.
80. Zhurkin,V.B., Tolstorukov,M.Y., Xu,F., Colasanti,A.V. and Olson,W.K. (2005) Sequence-dependent variality of B-DNA: an update on bending and curvature. In Ohyama,T. (ed.), *DNA Conformation and Transcription*. Georgetown TX/New York: Landes Biosci./Springer Sci. Bus. Media, pp. 18–34.
81. Khandelwal,G. and Bhyravabhotla,J. (2010) A phenomenological model for predicting melting temperatures of DNA Sequences. *PLoS ONE*, **5**, e12433.
82. Rando,O.J. and Ahmad,K. (2007) Rules and regulation in the *primary structure of* ch**ro**matin. *Curr. Opin. Cell Biol.*, **19**, 250–256.
83. Schnitzler,G.R. (2008) Control of nucleosome positions by DNA sequence and remodeling machines. *Cell Biochem. Biophys.*, **51**, 67–80.
84. Travers,A.A. and Klug,A. (1990) Bending of DNA in nucleoprotein complexes. In Cozzarelli,N.R. and Wang,J.C. (eds), *DNA Topology and Its Biological Effects*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 57–106.
85. Travers,A. and Drew,H. (1997) DNA recognition and nucleosome organization. *Biopolymers*, **44**, 423–433.
86. Lee,W., Tillo,D., Bray,N., Morse,R.H., Davis,R.W., Hughes,T.R. and Nislow,C. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.*, **39**, 1235–1244.
87. Mavrich,T.N., Ioshikhes,I.P., Venters,B.J., Jiang,C., Tomsho,L.P., Qi,J., Schuster,S.C., Albert,I. and Pugh,B.F. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.*, **18**, 1073–1083.
88. Haran,T.E. and Crothers,D.M. (1989) Cooperativity in A-tract structure and bending properties of composite TnAn blocks. *Biochemistry*, **28**, 2763–2767.
89. West,S.M., Rohs,R., Mann,R.S. and Honig,B. (2010) Electrostatic interactions between arginines and the minor groove in the nucleosome. *J Biomol Struct Dyn.*, **27**, 861–866.